

Robotic movement preferentially engages the action observation network

Emily S. Cross^{1,2*}, Roman Liepelt^{1,3}, Antonia F. de C. Hamilton⁴, Jim Parkinson¹,
Richard Ramsey⁴, Waltraud Stadler¹ & Wolfgang Prinz¹

¹Max Planck Institute for Human Cognitive & Brain Sciences, Department of Psychology, Stephan Straße 1a, 04103, Leipzig, Germany

²Radboud University, School of Psychology and Artificial Intelligence, PO Box 9104, 6500 HE, Nijmegen, The Netherlands

³Westf. Wilhelms-University, Department of Psychology, Fliednerstr. 21, 48149 Münster, Germany

⁴University of Nottingham, School of Psychology, University Park, NG7 2RD, Nottingham, England

* To whom correspondence should be addressed:

E. S. Cross
Radboud Universiteit Nijmegen
Department of Psychology
PO Box 9014
6500 HE Nijmegen, The Netherlands
ph: +31 24 361 00 81
email: e.cross@psych.ru.nl

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

Abstract

As humans, we gather a wide range of information about other people from watching them move. A network of parietal, premotor, and occipitotemporal regions within the human brain, termed the *action observation network* (AON), has been implicated in understanding others' actions by means of an automatic matching process that links observed and performed actions. Current views of the AON assume a matching process biased towards familiar actions; specifically, those performed by conspecifics and present in the observer's motor repertoire. In the present study, we test how this network responds to form and motion cues when observing natural human motion compared to rigid robotic-like motion across two independent functional neuroimaging experiments. In Experiment 1, we report the surprising finding that premotor, parietal, occipitotemporal regions respond more robustly to rigid, robot-like motion than natural human motion. In Experiment 2, we replicate and extend this finding by demonstrating that the same pattern of results emerges whether the agent is a human or a robot, which suggests the preferential response to robot-like motion is independent of the agent's form. These data challenge previous ideas about AON function by demonstrating that the core nodes of this network can be flexibly engaged by novel, unfamiliar actions performed by both human and non-human agents. As such, these findings suggest that the AON is sensitive to a broader range of action features beyond those that are simply familiar.

Keywords: action observation network, functional MRI, parietal, premotor, robots, dance

The human body in motion is a rich source of social data, conveying such information as the mover's identity, goals, intentions, and even thoughts, beliefs, or desires. Research into the social brain is advancing rapidly (Adolphs, 2010; Meltzoff, 2007), and a key component that helps to facilitate social interaction is believed to be an action observation network (AON; Gallese and Goldman, 1998; Gallese et al., 2004). This network, comprising premotor, parietal, and occipitotemporal cortices, is activated when observing other people in action (Cross et al., 2009b; Gazzola and Keysers, 2009; Grèzes and Decety, 2001).

Moreover, some components of this network, including inferior parietal and inferior premotor cortices, correspond to brain regions containing mirror neurons in non-human primates (Gallese, 2007; Gallese et al., 2004; Rizzolatti and Craighero, 2004). Recent neuroimaging work with humans shows that parietal and premotor nodes of the AON encode performed and observed actions within the same networks (Kilner et al., 2009; Oosterhof et al., 2010). It has been suggested that these parts of the AON permit 'direct' comprehension of other people's behavioral and mental states (Gallese and Goldman, 1998; Gallese et al., 2004). Furthermore, the discovery of neurophysiological processes linking action with perception has been taken as evidence in support of the ideomotor principle of action control. This prominent theory, introduced over 120 years ago, posits that actions are cognitively represented in terms of their sensory consequences, thus providing a common code for perception and action (James, 1890; Prinz, 1990).

The dominant model of the AON proposes that this network of brain regions responds most robustly when watching familiar, executable actions (Press, 2011). Brain regions associated with the AON show stronger responses

to humans than animals (Buccino et al., 2004), familiar than unfamiliar actions (Calvo-Merino et al., 2005; Cross et al., 2006; Cross et al., 2009a; Cross et al., 2009b), human agents than robotic agents (Chaminade et al., 2010; Miura et al., 2010; Shimada, 2010; Tai et al., 2004) or other non-human agents (Costantini et al., 2005; Engel et al., 2008), and same-race than different-race individuals (Avenanti et al., 2010; Liew et al., 2010). Moreover, both AON activity (Wheatley et al., 2007) and motor priming in behavioral tasks (Liepelt and Brass, 2010; Liepelt et al., 2010) are increased when participants believe they are observing another human, compared to a robotic or inanimate agent (although other studies report less clear effects of animacy instructions on AON activity, c. f. Stanley et al., 2010). Behavioral work investigating how observed actions influence simultaneously performed actions has reported greater interference effects when participants watch a human actor moving with human kinematics compared to moving with a constant velocity profile (Kilner et al., 2007a), and when watching a robotic agent moving in a human-like manner compared to a non-biological manner (Chaminade et al., 2005). These findings are consistent with the notion that observing actions with familiar kinematic features, which are within one's motor repertoire, results in greater AON activation than observing less familiar actions.

Findings concerning the impact of an agent's form on AON activity and motor resonance are also largely consistent with the familiarity hypothesis. A number of behavioral studies demonstrate a greater influence of action observation on action performance when actions are performed by human compared to non-human, symbolic agents, even when all movements have been precisely matched to display human kinematics (Brass et al., 2001; Gowen et al.,

2008). Such findings are further corroborated with evidence from electroencephalography (Oberman et al., 2005), magnetoencephalography (Kessler et al., 2006), near infrared spectroscopy (Shimada, 2010), and functional magnetic resonance imaging (fMRI; Chaminade et al., 2010; Miura et al., 2010), which all demonstrate that the AON is indeed more strongly engaged by human compared to non-human forms, even when all actions are matched for kinematics.

Based on these studies examining observation of familiar compared to less familiar form and motion cues, engagement of the AON in particular and the motor system in general is commonly taken as an indicator of spontaneous simulation of actions that are present in an observer's motor repertoire. Furthermore, some studies are now measuring engagement of AON in a form of reverse inference in which greater activity in these regions is taken as a sign that participants see the other actor as being 'like me', and lack of AON engagement is taken as a measure of failure to link self and other (Dapretto et al., 2006; Oberman et al., 2005).

In contrast to these results, emerging evidence suggests that the relationship between AON activity, the form of the observed agent and the motion of the observed action might be much more nuanced (Cross et al., 2009a; Gazzola et al., 2007; Ramsey and Hamilton, 2010). That is, the AON may not simply respond more when observing actions with familiar form or motion, a notion that departs from the familiarity hypothesis. One study found no difference in AON activity when participants observed a human hand and a robotic hand perform simple actions, such as grasping a cup (Gazzola et al., 2007). It is of interest that the robotic hand in this context differed from the

human hand on both form and motion parameters, thus making it impossible to know how these features independently influence responses within the AON. Another recent study asked participants to watch short animations of geometric shapes “retrieve” simple objects, such as a cookie or some keys (Ramsey and Hamilton, 2010). In this study, the shapes did not have human form or motion, but were animated to make them appear self-propelled and “alive”. The authors found activity in the parietal node of the AON when participants observed these abstract animations (Ramsey and Hamilton, 2010), similar to previous studies involving observation of human goal-directed action (Hamilton and Grafton, 2006).

As such, these latter studies provide preliminary evidence for a greater flexibility of information processing within the AON, which is not restricted to familiar action features. However, a key question remains regarding the relationship between observed form and motion within the AON. For example, in the studies that provide evidence for AON activation when observing non-human agents, motion parameters were not systematically manipulated (i.e., the robotic hand (Gazzola et al., 2007), animated shapes (Ramsey and Hamilton, 2010), and scrolling arrows (Cross et al., 2009a) all moved in just one manner, none of which resembled human motion). To this end, open questions remain concerning how this network processes unfamiliar, non-human motion cues independent of agent form, as well as how form and motion cues might interact.

In the present study, we systematically examine the relationship between motion and form cues in terms of action familiarity and activity within the AON. In the first experiment, we compare brain activity when observing a human dancer perform in a rigid, robotic manner, compared to a smooth, fluid, natural

manner. In the second experiment, we build upon the first by directly juxtaposing form and motion cues. We accomplish this by matching the movements of an articulated robotic figure with the human dancer's movements, in both the robotic dance and natural dance conditions. As such, the following experiments enable us to cleanly disambiguate the impact of motion (natural human vs. rigid robotic) and form (human vs. robot) cues on the AON.

Materials and Methods

Subjects. Twenty-two right-handed volunteers (12 men, 10 women; mean age 25.95 years, range: 20.5 – 33.6 years) participated in Experiment 1. Twenty-three different right-handed volunteers (10 men, 13 women; mean age 25.25 years; range 19.7 – 31.0 years) participated in Experiment 2.

Across both experiments, all participants were naïve to the purpose of the experiment, free of any neurological or psychiatric disorder, or were not on medication at the time of measurement. No participant reported having formal dance training, and while most participants reported occasionally dancing in clubs, no participant reported trying to “dance the robot”. All participants were strongly right-handed according to self-reported responses to the Edinburgh Handedness Inventory (Oldfield, 1971). All participants provided written informed consent and were monetarily compensated for their time, treated according to the ethical regulations laid out in the Declaration of Helsinki. The local ethics committee approved all experimental procedures.

Stimuli and Design.

Experiment 1: Thirty-six videos ranging in length from 7.7 to 9.8 seconds featured a professional break-dancer dancing in a natural, free-style manner (Video S1) or in a rigid, robotic manner, known as 'dancing the robot' (Video S2). Importantly, the videos were not altered in any way - the dancer was simply instructed to dance naturally and dance robotically, and the robotic dance videos used throughout both experiments feature this individual dancers' interpretation of dancing like a robot. The dancer performed to music that was of equivalent tempo across both dance styles. However, the videos used for the experiment did not include any audio information. In half the videos, the dancer wore a plain white mask with white gauze over the eyes, in order to render static all facial information. The motivation behind the mask manipulation was to determine whether robotic movements would be perceived as less human-like if access to facial information was obscured by a mask.

The 36 stimuli videos of Experiment 1 fell into a 2 x 2 factorial design, with factors face (levels: face visible; face masked) and movement (levels: natural; robotic - see Figure 4A). Of these videos, 8 featured the dancer dancing in a natural style with his face visible, 8 were of the dancer dancing in a natural style while wearing the white mask, 8 were of the dancer dancing in a robotic style with his face visible, and 8 were of the dancer dancing in a robotic style while wearing the white mask.

Experiment 2: Thirty-two videos ranging in length from 7.72 to 9.80 seconds were used. Of those videos, 16 were the same as those used in Experiment 1; 8 featuring the dancer moving in a natural manner (without the mask) and 8 featuring the dancer moving in a robotic manner (without the mask). The

remaining 16 videos were created with a Lego Bionicle™ action figure (model 7117, name: Gresh) and stop-motion animation, using Frame-by-Frame software (<http://web.mac.com/philipp.brendel/Software/FrameByFrame.html>). The videos were made by matching the Lego figure's limbs to the positions of the human dancer's limbs. This matching process was performed by overlaying real-time video of the Lego figure onto the prerecorded video of the human dancer. The original videos of the human dancer were advanced frame by frame, and the Lego figure's posture was adjusted to match the human's for each video frame. As the human videos were recorded at a rate of 25 frames per second, this resulted in a total of 193 to 245 static images of the Lego figure, which, when played back at the rate of 25 frames per second, precisely matched the human videos in duration. This resulted in 16 frame-matched videos featuring the Lego form: 8 with the Lego form moving in a natural human dance style (Video S3), and 8 with the Lego form moving in a robotic dance style (Video S4).

Motion Energy Quantification. Because videos within and across the robotic and natural dance style categories varied in the amount of limb motion, it was necessary to ensure that neural differences that were observed between movement styles were not due to more low-level action features, such as the number of movements or the size or scale of the movements in space. In order to do this, we quantified the motion energy in each video clip using a custom Matlab algorithm, based on work in motion recognition in computer science (Bobick, 1997). Such quantification of motion energy has been applied successfully before to stimuli used in neuroimaging studies of action observation (Schippers et al., 2010). With our particular algorithm, we converted each movie

to grey-scale, and then calculated a difference image between each pair of consecutive frames in each movie. The difference image was thresholded so that any pixel with more than 10 units luminance change was classified as 'moving'. The average number of moving pixels per frame and per movie was summed to give a motion energy score for that movie. Results from the motion energy quantification procedure are illustrated in Figure S1.

fMRI Task. During functional neuroimaging, all videos were presented via Psychophysics Toolbox 3 running under Matlab 7.2. The videos were presented in full color with a resolution of 480 x 270 pixels using a back projection system, which incorporated a LCD projector that projected onto a screen placed behind the magnet. The screen was reflected on a mirror installed above participants' eyes. In Experiment 1, participants completed 36 trials from each category organized in a pseudorandom design, with 36 probe trials in total (9 probe trials from each category). In Experiment 2, participants completed 32 trials from each category organized in a pseudorandom design, with 32 probe trials in total (8 probe trials from each category). In both studies, the intertrial intervals were pseudologarithmically distributed and ranged between 7 – 9 seconds.

Across both experiments, participants received identical instructions. Their task was to watch each video closely, and to stay alert for probe trials. Probe trials occurred after 25% of experimental trials, and were signaled by a short display of a question mark, followed by a 2-s. video clip that participants were required to decide whether this video clip was part of the larger video they had previously viewed, or not (Figure 1). Probe clips were always taken from the same category as the full experimental videos (i.e., if a probe trial followed an

experimental trial where the human dancer danced like a robot, then the 2-s. test clip would also feature a human dancer dancing like a robot). Probe trials were designed so that 50% of trials were correct matches, and 50% were incorrect matches.

fMRI Data Acquisition. Data acquisition for both experiments was conducted at the Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany). Functional images were acquired on a Bruker 3-T Medspec 20/100 whole-body MR scanning system, equipped with a standard birdcage head coil. Functional images were acquired with a single shot gradient echo-planar imaging (EPI) sequence with the following parameters: echo time TE = 30 ms, flip angle 90°, repetition time TR = 2,000 ms, acquisition bandwidth 100 kHz. Twenty-four axial slices allowing for full-brain coverage were acquired in ascending order (pixel matrix = 64 x 64, FOV = 19.2 cm, resulting in an in-plane resolution of 3 mm x 3 mm, slice thickness = 4 mm, interslice gap = 1 mm). Slices were oriented parallel to the bicommissural plane (AC-PC line). Geometric distortions were characterized by a B0 field-map scan (consisting of a gradient-echo readout (32 echoes, inter-echo time 0.64 ms) with a standard 2D phase encoding). The B0 field was obtained by a linear fit to the unwrapped phases of all odd echoes. Prior to the functional run, 24 two-dimensional anatomical images (256 x 256 pixel matrix, T1-weighted MDEFT sequence) were obtained for normalization purposes. In addition, for each subject a sagittal T1-weighted anatomical scan high-resolution anatomical scan was recorded in a separate session on a different scanner (3T Siemens Trio, 160 slices, 1 mm thickness). The

anatomical images were used to align the functional data slices with a 3D stereotactic coordinate reference system.

Scanning parameters were held constant across Experiments 1 and 2, with only one change: in a single fMRI run, a total of 1,293 images were collected for Experiment 1, and a total of 1,150 images were collected for Experiment 2. In addition, following the fMRI portion of Experiment 2 only, participants were asked to complete a short dance rating survey. For this survey, participants watched each video once more, and rated on a 1-7 scale their perceived ability to reproduce each movement, with 1 corresponding to “I could not come close to reproducing that dance right now” and 7 corresponding to “I could reproduce that dance perfectly right now”. Similar scales have been used in the past, and validated for ratings of perceived ability tracking with actual physical ability (Cross et al., 2006).

fMRI Data Analysis. Data were realigned, unwarped, corrected for slice timing, normalized to individual participants’ T1-segmented anatomical scans with a resolution of 3 mm x 3 mm x 3 mm, and spatially smoothed (8 mm) using SPM8 software. A design matrix was fitted for each participant with regressors for each of the four video types in the factorial design, as well as a parametric modulator for each regressor that expressed the mean motion energy of each video. The inclusion of these parametric regressors should model out any differences between conditions that are due simply to the different amounts of movement between conditions. One additional regressor was included that encompassed the probe video and question phase. Each trial was modeled as a

boxcar with the duration of that video convolved with the standard hemodynamic response function.

Experiment 1: The first group-level analysis evaluated which brain regions were more active when watching the dancer perform in either a robotic or natural dance style, compared to baseline (in this instance, our baseline is an implicit measure, as we did not explicitly model fixation onsets and durations). This was achieved by collapsing across the ‘face/mask’ factor, and comparing all robotic dance to an implicit baseline measure, and comparing all natural dance to the same implicit baseline measure. The main effect of movement style was calculated in a random effects analysis. In addition, the main effect of face visibility was calculated, as were the interactions between these two factors. These latter two analyses are not the focus of the present paper, but the results from each are presented in Table I for completeness. All contrasts were evaluated at the $p_{\text{uncorrected}} < 0.001$, $k = 10$ voxel threshold, and only results that reached a significance level of $p < 0.05$, FDR-corrected, are discussed in the main text.

Experiment 2: Both main effects and interaction analyses were evaluated at the random effects level. For the main effect of movement style, comparisons were collapsed across human and Lego forms, and for the main effect of form, comparisons were collapsed across natural and robotic dance styles. Findings from all analyses are reported in Table II.

Results

Behavioural Results

Figure 2 illustrates that participants performed the task at a high level of accuracy across Experiment 1 (88.0%) and Experiment 2 (89.1%). In more detail, no main effects for dance style or face presence emerged in Experiment 1 (both p -values > 0.05), but there was an interaction between dance style and face presence, which suggests participants' performed marginally less accurately when performing the matching task for trials where the face was unmasked and the dancer danced in a robotic manner, compared to a natural manner, $F_{1,21} = 7.47$, $p = 0.012$. Statistical analysis of behavioral data from Experiment 2 revealed no main effects or interactions from the accuracy data. Together, these behavioral data suggest that participants were carefully attending to all video categories.

Neuroimaging Results

In Experiment 1, a random effects analysis ($p < 0.05$ whole brain cluster-corrected) demonstrated that compared to an implicit baseline, observing natural dancing activated frontal, parietal, and occipitotemporal brain regions. Compared to the same baseline, observation of robotic dancing activated the same regions (Figure 3A; 3B). Direct comparison of the two kinds of dancing revealed striking differences: observation of robotic dancing engages parietal, premotor and middle temporal regions more than observation of natural dancing (Figure 3C; Table I). The inverse contrast demonstrated greater recruitment of visual cortex, centered on the lingual gyrus (Figure 3D; Table I).

In Experiment 2, we sought to determine whether increased activation in the AON reflects a sensitivity to form and motion mismatches (i.e., a human form

does not normally move like a robot; Figure 4A). As in Experiment 1, we found that observation of robotic compared to natural motion engaged premotor, inferior parietal and middle/superior temporal cortices, while again only visual regions showed selectivity for natural motion compared to robotic motion (Figure 4B; 4C; Table II). Observation of the Lego figure compared to the human figure lead to bilateral ventral occipitotemporal and left inferior parietal activation (Figure 4D). The inverse contrast did not yield significant clusters. The interaction between watching a human dance naturally and a Lego figure dance robotically, compared to watching a human dance robotically and a Lego figure dance like a human revealed one cluster of activation within the left temporoparietal junction (TPJ) region (Figure 5; Table II). Thus, it would appear that this region responds to typical form-action pairings more than atypical ones. The inverse interaction revealed greater activity only within the left middle occipital gyrus, negating the possibility that AON activity seen in this present study is simply due to unusual agent and action pairings.

An additional fMRI analysis was performed to evaluate the simple effect of robotic dance > natural dance only for videos with the Lego form. This analysis was performed in order to rule out the possibility that the strong effects observed when robotic dance movements are contrasted with natural dance movements are driven exclusively by the human form moving in an uncharacteristic manner. The results from this analysis confirm that this is not the case (Figure S2). Simply watching an articulated Lego form move in a robotic manner compared to a natural human manner engages the AON.

Post-scan behavioral ratings of how well participants believed they could perform the dances showed that robotic dance was rated as more difficult to

reproduce than natural dance (Figure 6). When a 2 x 2 repeated measures ANOVA was performed on participants' self-report scores, a main effect of motion emerged, with participants rating the robotic dance sequences as significantly more difficult to reproduce ($M = 4.61$; $SD = 0.78$) than the natural dance sequences ($M = 4.13$; $SD = 0.89$; $F_{1,22} = 9.458$, $p = 0.006$). There was no main effect for form ($F_{1,22} = 2.88$, $p = 0.104$), nor was there an interaction between form and motion ($F_{1,22} = 0.13$, $p = 0.721$).

Discussion

Across two independent studies with 45 participants, we demonstrate that human inferior parietal, premotor, and occipitotemporal cortices respond more robustly to rigid, robot-like motion than to natural human-like motion for both human and Lego robot forms. These results are challenging to reconcile with previous studies suggesting that the AON is preferentially responsive to human agents (Chaminade et al., 2010; Costantini et al., 2005; Engel et al., 2008; Shimada, 2010; Tai et al., 2004) or an observer's prior motor experience (Aglioti et al., 2008; Calvo-Merino et al., 2005; Cross et al., 2006; 2009a). They suggest that the hypothesis that the AON responds preferentially to actions that are familiar does not fully capture the sensitivity of this neural system, thus creating a need for an updated model of AON function. In the following discussion, we consider several interpretations of the current data and put forward an account of AON function that can accommodate the current findings with seemingly conflicting results from previous studies. We also suggest avenues for future research that could test these proposals.

Effects of Agent Motion

In the first experiment, the comparisons to baseline revealed robust AON activity both when participants watched a human figure dance in a normal, fluid human manner or in a rigid, robotic manner. This unsurprising result confirms that parietal and premotor regions, as well as parts of visual cortex, are active when watching a human in motion. However, when observation of human-like and robot-like motion is directly compared, large clusters within bilateral inferior parietal lobules, occipitotemporal cortices, and right ventral premotor cortex respond more strongly to the rigid, robot-like dance movements. These findings show that observing a real human move in a robotic manner lead to increased activation of the AON, compared to watching the same agent move in a more human-like manner.

These data run contrary to prior work on observation of virtual human avatars moving with robotic kinematics (Shimada, 2010), or robotic model dressed to look like a human arm (Tai et al., 2004). Our results are also the opposite of that predicted by the familiarity hypothesis of AON function. Here we see that the *less* human-like and *less* familiar the actions appeared, the stronger the AON responded. No AON regions showed the opposite pattern of a preference for human-like compared to robot-like motion. Instead, activation within the lingual and middle occipital gyri was found when observing natural human compared to robotic dance motion. This could be driven by concurrent movement of multiple limbs in opposing directions during the natural, human motion condition (as opposed to the serial, one limb at a time, movements of the robotic motion condition), consistent with work showing this region of visual

cortex to be responsive to motion moving in multiple directions (Van Oostende et al., 1997).

The findings from Experiment 2 replicate and extend those from Experiment 1. The inclusion of both human and robotic forms in Experiment 2 enabled us to directly test how form and motion cues contribute independently and interact to influence AON function. Here again, we report stronger AON activation when participants watched the robot-like motion compared to the natural human motion, independent of whether the agent was a human or a Lego robot. This form-independent increased activation of the AON when observing robot-like motion suggests that motion is more critical than form in driving AON activity in this instance. Additional analyses of Experiment 2's data lend further weight to our suggestion that the AON is not strictly biased to respond to action features that are familiar. The simple effects analysis evaluating robotic vs. natural human movement styles for the Lego form only demonstrates that observation of agents that neither look nor move like the observers leads to robust activation across parietal, premotor, and occipitotemporal cortices. Finally, the post-scanning behavioral follow-up test, which assessed participants' perceived ability to physically perform the observed actions, demonstrates that participants found the robotic actions (as performed by either the human or Lego agent) to be more difficult to reproduce than the human actions. As such, these data provide further evidence in contradiction to past work on perceived action ability and AON activation (Calvo-Merino et al., 2005; Cross et al., 2006), by demonstrating a dissociation between how well participants think they can perform an action and activation of the AON.

Previously, the better participants thought they could reproduce a dance sequence, the higher the BOLD signal was in left parietal and ventral premotor cortices (Cross et al., 2006). In the present study, the data support the inverse relationship. In other words, when the participants from Experiment 2 were interrogated about their ability to reproduce the observed movements on the same scale used previously (Cross et al., 2006), they rated the robotic movements as significantly more difficult to reproduce than the human-like movements (Figure 6). Thus, a clear dissociation emerged in the present study between general performance ability and AON activation. That is, the AON responded more to the perception of actions that were perceived as more difficult for participants to reproduce compared to actions that were perceived as easier to reproduce. As such, when considered with previous findings, our data hint that the relationship between perceived performance ability (or familiarity in general) of an observed action and AON activity is non-linear. Greater BOLD signal can be associated with unfamiliar actions, which are not frequently performed by individuals (as demonstrated in the current study), as well as familiar actions that are frequently performed by individuals (Calvo-Merino et al., 2005; Cross et al., 2006). In the final part of the discussion, we outline a hypothesized account of AON function that could possibly explain these disparate findings.

Effects of Agent Form

Occipitotemporal and inferior parietal cortices showed a greater response when observing the Lego robot compared to the human agent dancing, independent of movement style. Increased activation of ventral

occipitotemporal regions when observing the robotic form is consistent with the notion that a non-human agent with a humanoid body recruits visual regions implicated in body processing (Chaminade et al., 2010; Peelen and Downing, 2007). One possibility is that this activity reflects increased demands to visually compare the observer's body with the observed body, a notion consistent with prior work on body perception (e.g., Felician et al., 2009). Furthermore, imagined transformations of the human body (such as those that might occur when implicitly comparing one's own body with the observed Lego figure's body) have been reported within similar regions of the parietal cortex (Jackson et al., 2006; Zacks et al., 2002). Another factor that might contribute to greater occipitotemporal activity when viewing the Lego robot compared to the human is the greater motion energy contained in the videos featuring the robotic agent (Figure S1). Even though we were careful to model out mean motion energy from our experimental design, we calculated only a single value per video, which means that not every feature was perfectly equated across the duration of each video stimulus. The higher motion energy level for the Lego robot is attributable to the fact that this figure has relatively longer arms and legs than the human, which results in more pixels being displaced when the it "performs" the same movements. Prior evidence supports the notion that portions of the lateral occipital cortex track with increasing optic flow contained in visual stimuli (Beer et al., 2009), similar to that seen in the Lego robot compared to human stimuli in the present study.

The interaction analyses shed light on how the brain processes consistent and contrasting form and motion pairings. The first interaction evaluated brain regions most active to typical agent-action pairings (i.e., human form dancing in

natural human style and Lego form dancing in a robotic style) vs. atypical agent-action pairings (i.e., human dancing robotically and Lego form dancing in a natural human style). The activation of left TPJ is consistent with prior work that implicates this region as part of multimodal association cortex (Decety and Sommerville, 2003). Its emergence from this contrast might signal its responsiveness to the congruence between agent and action. The finding of increased activity within the middle occipital gyrus for the inverse interaction suggests that this region may be recruited to perform increased visual analysis of atypical agent/action pairings, such as when a human dances in a robotic manner or a Lego form moves in natural, human manner (Chaminade et al., 2010).

Why Might the AON Respond Most Strongly When Watching Robotic Actions?

a. Low-level visual features

There are several possible explanations for the present results. First, distinct brain responses to human and robotic dance could be caused by differences between low-level motion features such as smoothness or the amount of movement in each stimulus. Because quantitative differences exist between the movements seen in the natural human dance and robotic dance conditions, the visual inputs for both conditions are different. This could then result in distinct patterns of AON engagement, which have been driven in a bottom-up manner from visual cortical input. We suggest this explanation is unlikely for several reasons. The first and most important reason is because we explicitly quantified and modeled out differences in motion energy between all conditions in our design. Secondly, our movement quantification algorithm

revealed other features about the stimuli that are relevant to consider if a motion energy-based account is to be adopted. One is that there is more, not less, motion in movement sequences featuring the natural human style of dance (Figure S1). Therefore, our finding of greater AON activation when observing action sequences with *less* overall movement is counterintuitive if a bottom-up visual input explanation is adopted.

Another feature that may differ between conditions is implied muscle tension when perceiving natural and robotic dance styles. It is likely that execution of the robotic style movements required greater muscle tension compared to performing the natural human style movements. Furthermore, observing actions involving greater muscle activity leads to greater excitability of primary motor cortex (Alaerts et al., 2010; Obhi and Hogeveen, 2010). However, we believe an explanation of our data *only* in terms of muscle tension differences is unlikely. First, such an explanation should be able to account only for differences during observation of a human dancer moving in a natural or a robotic style, since the Lego figure used in Experiment 2 has no muscles and thus no differences in muscle tension between dance styles. By this logic, if differences in perceived muscle tension were responsible for our findings, we should see no response in the AON for the simple effect of robotic versus natural human dance style when evaluating the effects of watching the Lego figure only (Figure S2). As is clear from this figure, even when an agent with no muscles is observed, greater activity within AON regions still emerges when watching the robotic compared to natural movement style. While it is conceivable that participants inferred the activity of non-existent muscles in the Lego robot, this possibility would need to be tested explicitly before such an explanation could be

supported. In sum, we acknowledge that differences in perceived muscle tension when observing the human agent, and even inferred muscle tension when watching the Lego figure, could partially underpin the differences in AON activity seen when observing robotic compared to natural human movement styles. However, due to a paucity of research investigating how perception of actual and inferred muscle tension is processed in the human brain, we cannot make strong claims regarding this interpretation. Instead, we encourage further investigation into the relationship between perceived and inferred levels of muscle tension during action observation.

b. Experimental task

A second possible explanation relates to the experimental task which participants were engaged in. In the vast majority of action observation studies (Buccino et al., 2004; Calvo-Merino et al., 2005; Cross et al., 2009a,b; Gazzola et al., 2007; Miura et al., 2010), participants simply observe stimuli or sometimes judge whether they can perform the action themselves (Cross et al., 2006). In the present study, participants were engaged in an explicit memory / encoding task, which might in and of itself demand deeper engagement of the AON, as suggested previously (Zentgraf et al., 2005). Put another way, when participants observe sequences with the intent of remembering the movement content, if the movements are unfamiliar compared to familiar, this could lead to stronger AON activity as these brain regions work harder to create an action representation where very little prior information is available. Such an interpretation is also consistent with a prediction error account of the data (discussed in more detail below). It is of note, however, that memory performance did not differ with trial type (Figure 2), so there is no behavioural evidence that the memory task was

more difficult when observing robotic movements compared to human-like movements. Ongoing research in our laboratory is investigating the relationship between distinct patterns of BOLD activity within the AON across conditions whilst task performance remains equivalent.

c. Attention

A third possible explanation of the present results could be differences in attention when observing the natural compared to robotic movement styles. It is possible that participants found the robotic dance style more 'engaging' than the human dance style, and this difference in engagement drove the differences in BOLD signal. Since attention is a broad construct with many different component parts, we consider two possible attentional accounts, first in terms of contextual incongruency, and second in terms of salience of the observed actions in relation to one's motor capabilities.

First, could contextual incongruency between seeing people dance like robots or robots dance like humans be driving the response in the AON? This question was assessed through calculation of the statistical interaction of form and motion. One brain region, right middle occipital gyrus, was sensitive to unusual action-agent pairings (i.e., the human dancing as a robot and the robot dancing as a human; see Figure 5). If the AON were sensitive to unusual action-agent pairings, we would expect responses in the AON for this interaction analysis. As we do not find such a response, it is difficult to explain our findings in terms of a contextual incongruency account. Although previous fMRI studies have shown that medial prefrontal cortex, TPJ and middle temporal gyrus are sensitive to contextual incongruency, in terms of how 'rational' actions appear given a certain context (Brass et al., 2007; Jastorff et al., 2011; Liepelt et al., 2008;

Marsh and Hamilton, 2011), we show that the AON is not sensitive to incongruent combinations of form and motion cues. Instead we show that incongruent action-agent pairings are encoded in middle occipital brain regions.

A second attention explanation relates to the salience of the observed movements. In particular it is possible that actions that are not in one's own motor repertoire (like robotic actions used in the current study) are more salient to individuals and capture attention more than actions we see and perform more frequently (like natural human motion). In order to reconcile the present data with prior findings, it could also be that when we are particularly skilled with certain actions, such as when professional dancers learn to perform a choreographed sequence of dance moves (Cross et al., 2006), these actions are more salient than actions with which we have not had such extensive physical experience. This explanation posits salience as a mediator of the BOLD effect, but the factor driving the salience of the action is the participant's motor familiarity with that action. We expand upon this explanation when hypothesizing a relationship between BOLD response and action familiarity below, which can account for disparate prior results.

In sum, there is no single, unambiguous attention-based explanation of our data, and any such explanation would leave unanswered the question of what drives greater attention to one stimulus set compared to another. While there is evidence that explicitly directing attention to different features of an action modulates the AON response (Chong et al., 2008; de Lange et al., 2008; Spunt et al., 2011), there is no empirical evidence that our stimuli demand systematically different levels of attention, nor why they might do so. However, while we suggest that a difference in attention for the two movement styles is

not an explanation for the present findings in and of itself, attention could nonetheless be mediating the pattern of results we report here. Continued investigation into how brain systems for attention and action observation interact would be valuable.

d. Prediction error

A fourth possible explanation of the present results could be greater engagement of compensatory top-down modulation of the AON (Schubotz, 2007) or greater prediction error (Neal and Kilner, 2010) when observing actions with robotic motion compared to natural human motion. The predictive coding account of the AON (Kilner et al., 2007b, c; Neal and Kilner, 2010) is based on empirical Bayes inference. It posits that the AON functions to minimize prediction error through reciprocal interactions among levels of the cortical hierarchy (parietal, premotor, and superior temporal regions; Kilner et al., 2007c). When observing an agent moving, we have prior expectations about how they might move based on the agent's goal (Gallese and Goldman, 1998), the environmental context (Liepelt et al., 2009), or the form of their body (Buccino et al., 2004). These prior expectations have an associated standard deviation, which signifies a level of confidence in each prior. The comparison between the predicted movements of the observed agent (based on the observer's action system) and the observed motion generates a prediction error. In the present study, when an observer has few or no action priors from either visual or physical experience (such as in the robotic motion conditions), prediction error is greatly increased, compared to when they have considerably more experience (such as in the natural human motion conditions). This discrepancy in prediction error between conditions would result in greater AON activation.

Such activation might reflect increased demands to learn, predict, or otherwise assimilate atypical actions into a more familiar biological motion template.

Integration of Prior and Present Findings

In the previous section, we outlined four possible accounts that might explain the present findings. As we mentioned, the present data contradict findings from prior experiments that examined physical or visual experience and action perception (Calvo-Merino et al., 2005; Cross et al., 2006; Cross et al., 2009b), and any one of the four accounts offered above (or a combination of them) might underlie the discrepancies observed between previous and present findings. As such, we are left with the challenge of explaining why prior work frequently demonstrates stronger AON activity when viewing familiar compared to unfamiliar movements (e.g., Buccino et al., 2004; Calvo-Merino et al., 2005; Cross et al., 2006; Shimada, 2010) and the present two experiments demonstrate the opposite pattern.

One possibility is that if differences in muscle tension explain the discrepancies between past and present findings, this would be a factor orthogonal to familiarity. Considering this possibility further, it is reasonable to think that the rigid, mechanical postures and 90° positioning of the limbs and torso in relation to each other in the present robotic motion stimuli could be the driving force behind higher AON activity when viewing robotic compared to natural human-like movement. The reason such differences might have been absent in prior studies is because in many of these studies (e.g., Gazzola et al., 2007; Kilner et al., 2003; 2007a; Tai et al., 2004), the natural and robotic stimuli moved according to roughly the same trajectories, with variations emerging mainly between velocity profiles (and not actual or inferred muscle tension).

Thus, while further work would be required to support or refute this explanation for differences between prior and present findings, it is possible that muscle tension variations are the main factor that account for the different pattern of findings we report here.

We argue, however, that the discrepancy between past and present findings is more parsimoniously explained by differences in familiarity. Specifically, the key difference between the present study and past studies of action familiarity (Buccino et al., 2004; Calvo-Merino et al., 2005; Cross et al., 2006; 2009b; Shimada, 2010) is the range of familiarities tested. Some studies compared extensively rehearsed actions to actions one might typically see and perform, but have not been explicitly rehearsed (Calvo-Merino et al., 2005; Cross et al., 2006; Cross et al., 2009b). In contrast, the current study compares unfamiliar actions, which individuals had little or no experience with, to more familiar actions one might typically see and perform. Across these studies, therefore, it might be possible to conceive of a familiarity continuum that ranges from 'unfamiliar' to 'generally familiar' to 'extensively familiar'. When considered in conjunction with BOLD responses in the AON, one possibility is that the relationship between action familiarity and AON response is not linear. Rather, it could be the case that strong BOLD signal can emerge for both highly unfamiliar actions and highly familiar actions, with weaker BOLD signal for 'generally familiar' actions (Figure 7). What is needed at this stage is research that systematically investigates this proposed spectrum of action familiarity to test this hypothesis.

Finally, it is worth considering that this proposed non-linear relationship between BOLD signal and familiarity is potentially compatible with

the Bayesian model of action observation described above (Kilner et al., 2007b; c). Observation of unfamiliar actions should lead to increased prediction error and greater BOLD signal, compared to observation of ‘generally familiar’ actions. In contrast, observation of extensively familiar actions where participants can make very specific predictions about how an action will continue based on prior experience might *also* lead to a higher prediction error compared to observation of ‘generally familiar’ actions, if the confidence in an action prior is high and the sensory input deviates from this. Such a proposal is still speculative at this stage, and there are undoubtedly a number of other possible ways in which BOLD signal within the AON and action familiarity may interact. More detailed computational modeling of how prediction error is modulated under different degrees of action familiarity would be useful to determine whether this proposed integration of prior and present results is valid.

Conclusion

Overall, our data demonstrate clear evidence that the action observation network responds more robustly to robotic motion cues compared to natural human motion cues, independent of whether the form of the agent is human or robotic. This finding suggests that action features that are unfamiliar to individuals compared to familiar can preferentially engage the AON. This result deviates from the dominant view that the perception of familiar action features produces a greater response in the action observation network. We have considered a variety of possible explanations for this finding, and suggest that there may be a non-linear relationship between BOLD signal and action familiarity. That is, heightened BOLD signal can be associated with both highly

unfamiliar actions and highly familiar actions compared to actions that are at neither end of a familiarity continuum. This means that activation of the AON cannot be taken as an indicator that a participant is physically or visually familiar with the observed action, or necessarily as evidence in support of the familiarity hypothesis. In other words, interpreting engagement of AON activity only as an indicator of social closeness is not feasible. Other factors, which might include novelty, task learning, and predictability can also substantially impact the BOLD signal. Precise articulation of what those factors are and how they might be modulated presents an intriguing challenge for future research.

ACKNOWLEDGEMENTS

We thank Carmen Hause for her invaluable help with the dance filming and stop-motion animations, M. Senf for his robotic dancing expertise, Andreas Wutz, Rafael Schultz-Kraft, Jana Strakova and Louise Kirsch for assistance implementing the experiments, and two anonymous reviewers for helpful suggestions on previous drafts of this manuscript. This research was supported by a postdoctoral research fellowship from the Alexander von Humboldt Foundation to E.S.C.

Figures

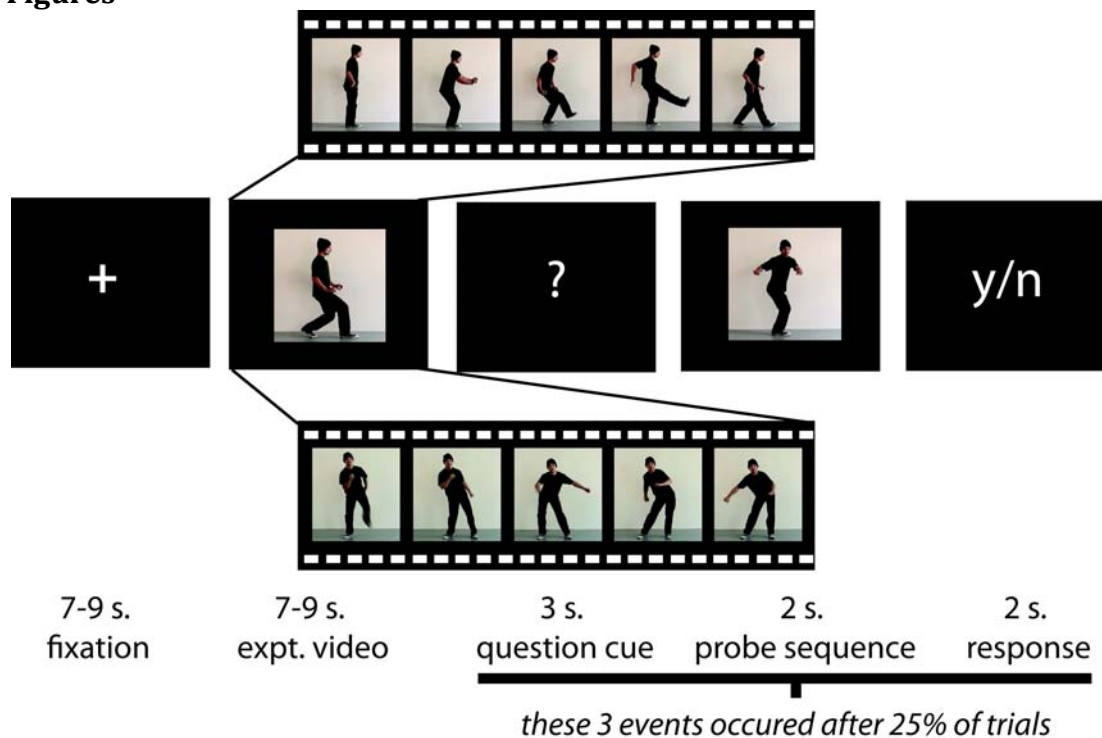


Figure 1. Representative experimental stimuli and timecourse (Experiments 1 and 2). The study began with a fixation cross, which was followed by a dance video 7-9 seconds in length. After 25% of the trials, a ‘?’ appeared, signaling to participants that they would next see a 2-second probe video. Participants’ task was to decide whether this video segment was part of the longer video they had just watched. Probe sequences were always chosen from the same category as experimental videos (i.e., a face-visible natural dance style probe video would always follow a face-visible natural dance style experimental video, and so on). Following the probe video, participants had 2 seconds to respond whether the probe video was or was not part of the previously observed experimental video.

In Experiment 1, the videos featured a human dancing in a robotic style (top video segment) or in a natural human style (bottom video segment), and the dancer’s face was either exposed (shown) or covered with a white mask (not shown). In Experiment 2, the videos featured a human dancing in a robotic style (top video segment) or in a natural human style (bottom video segment), or a Lego figure dancing in a robotic or natural human style (not shown).

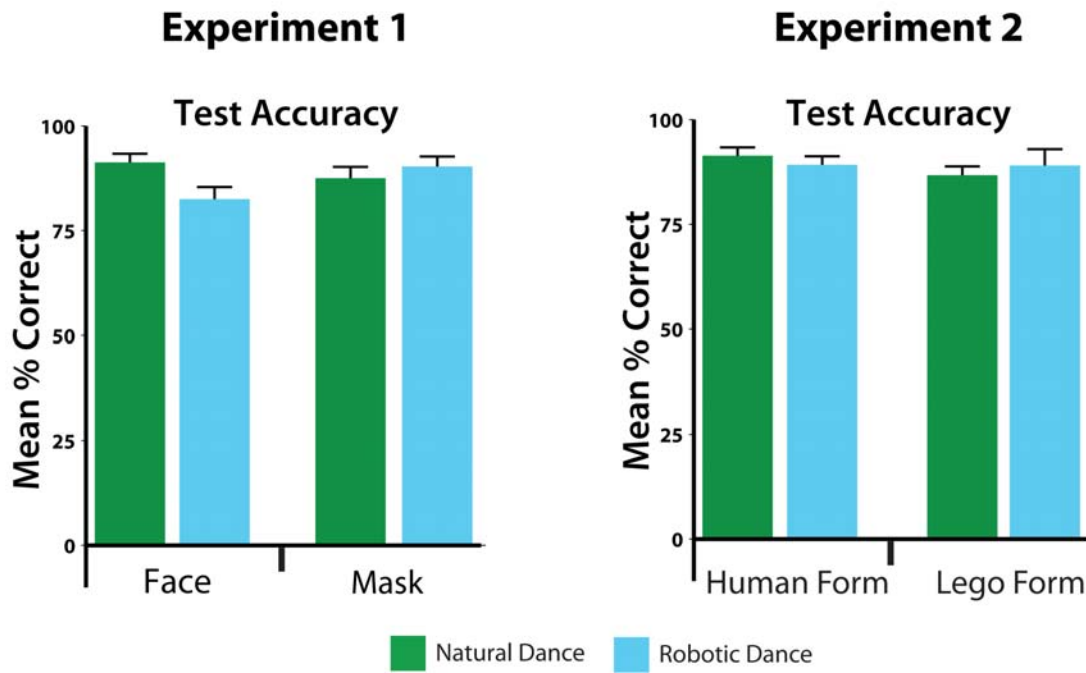


Figure 2. Behavioral data from fMRI task (Experiments 1 and 2). Plots illustrate mean accuracy (expressed as percent correct). Participants from Experiment 1 demonstrated an interaction between face presence and dance style, manifest as less accurate performance when observing a person with an exposed face dance robotically compared to dancing naturally. No other main effects or interactions were observed in Experiment 1 or Experiment 2.

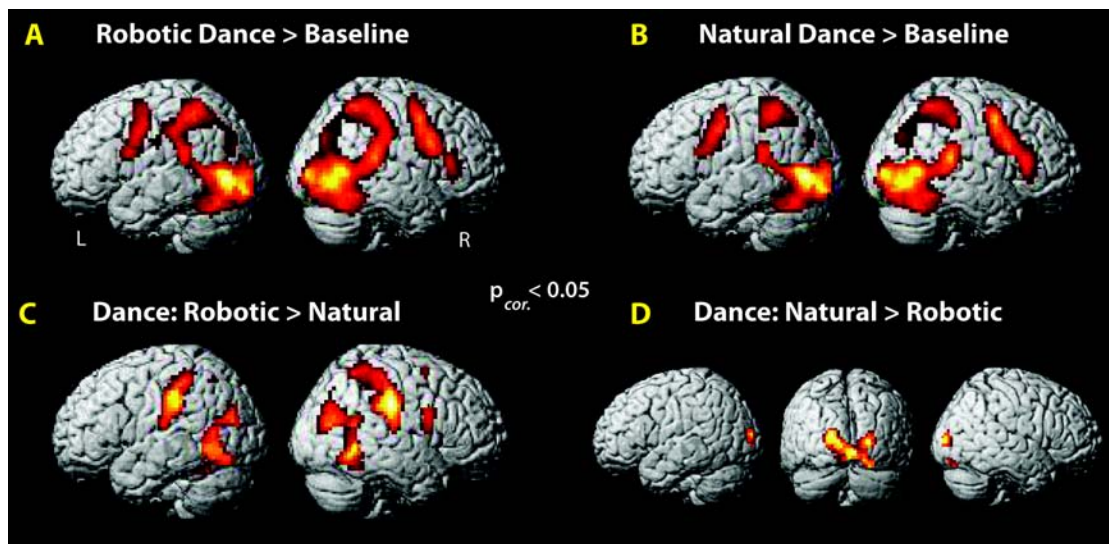


Figure 3. Main effects of motion manipulation, Experiment 1. Watching robotic (A) or human (B) dancing compared to baseline broadly activated the AON. Direct comparisons revealed stronger AON activation when watching robotic dancing (C), and stronger middle occipital gyrus in the inverse contrast (D; $p_{corrected} < 0.05$; Table 1).

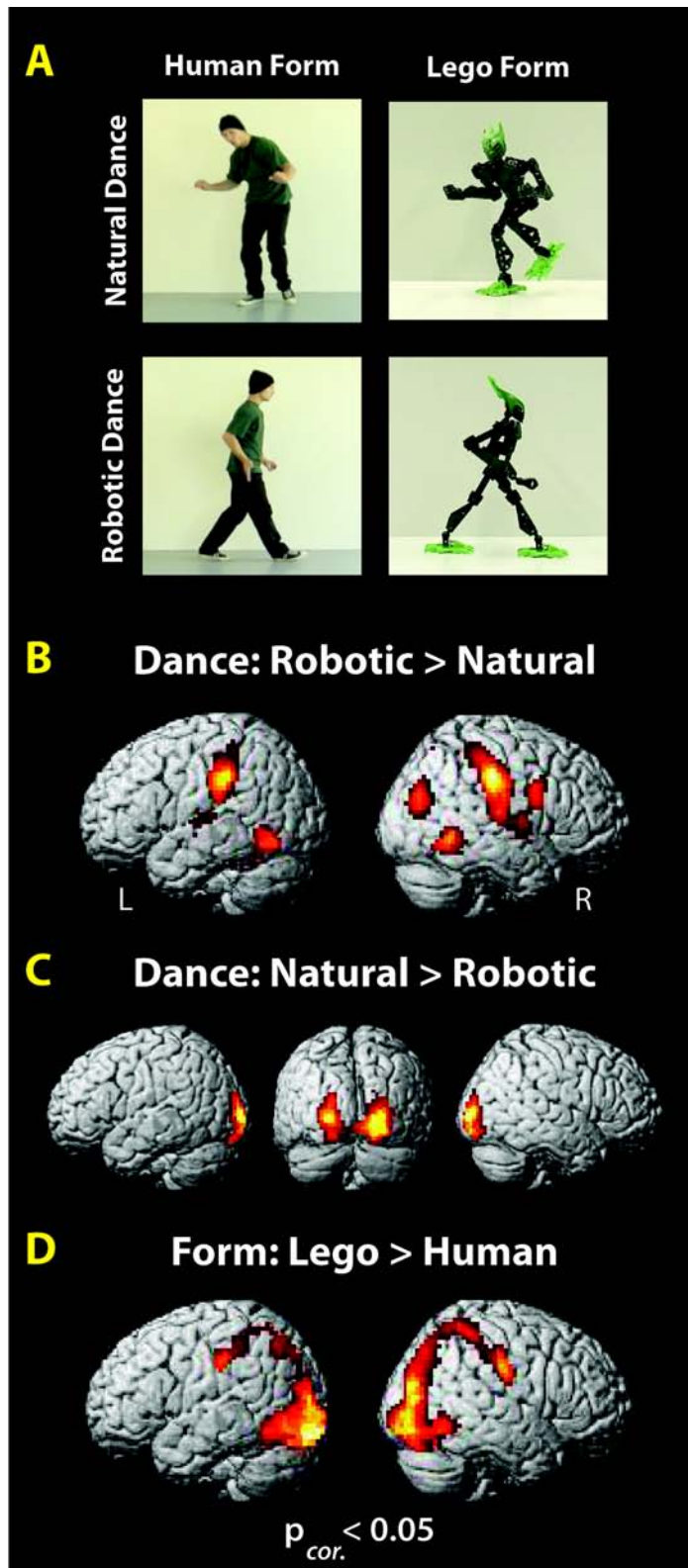


Figure 4. Experimental design and main effects from Experiment 2. The 2 x 2 factorial experimental design (A) and brain regions showing a greater BOLD response for robotic (B) and human motion (C) and robotic form (D; $p_{corrected} < 0.05$; Table 2).

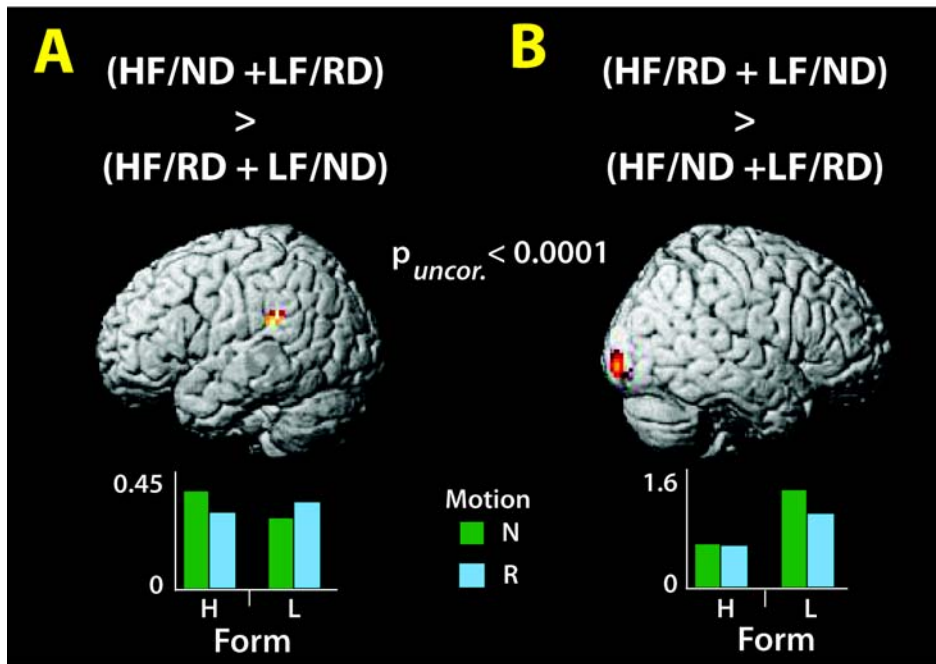


Figure 5. Interactions between form and motion (Experiment 2). (A) One cluster emerged within the left temporoparietal junction when watching a human dance normally and a Lego form dance robotically, compared to a human dance robotically and a Lego form dance naturally as a human would. (B) The inverse interaction revealed one cluster within right middle occipital gyrus, which responded most robustly to the Lego form dancing naturally. Abbreviations: HF = human form; ND = normal dance; LF = Lego form; RD = robotic dance ($p_{uncorrected} < 0.001$; Table 2).

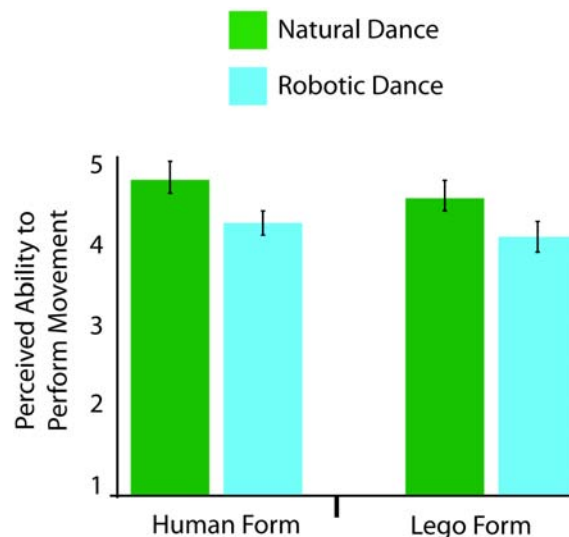


Figure 6. Participants' perceived ability to reproduce the dance sequences (Experiment 2). When participants were asked to rate their ability to reproduce each dance sequence on a 1-7 scale (anchors: 1 = not possible to reproduce at all; 7 = could reproduce perfectly), they rated the robotic dance sequences as significantly more difficult to reproduce than the natural dance sequences. No main effect of form emerged, nor did an interaction between form and motion (both p 's > 0.10).

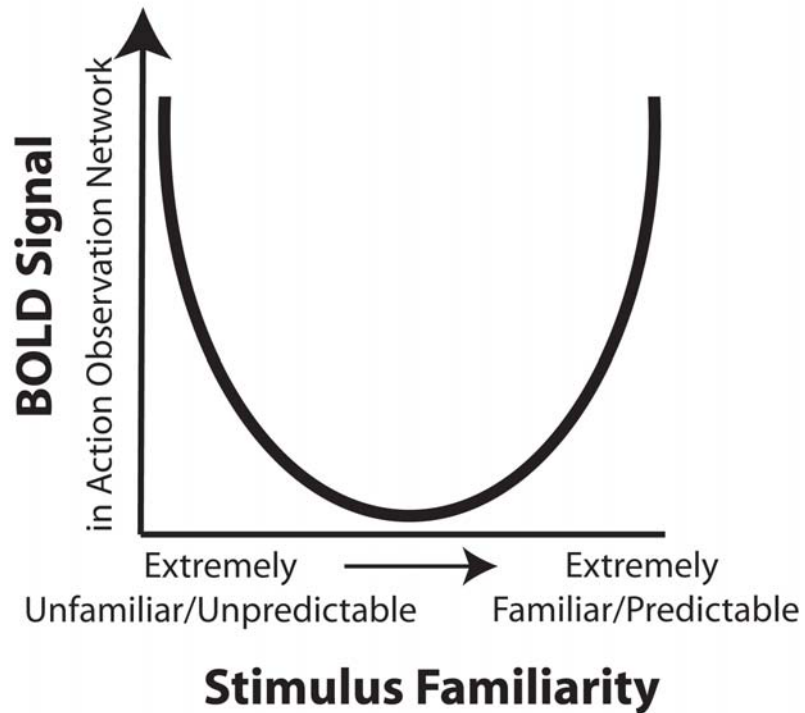


Figure 7. Hypothesized relationship between BOLD response and action familiarity. One way in which the present findings might be integrated with findings from many previous studies on familiarity and AON function is if strong BOLD signal is associated with both highly unfamiliar actions and highly familiar actions, with weaker BOLD signal for ‘generally familiar’ actions. At present, this interpretation is highly speculative and will require thorough testing to validate. However, such a relationship is consistent with a predictive coding account of AON activity (see main text).

TABLE 1

Anatomical Region	BA	MNI Coordinates			Putative Funct. Name	<i>t</i> - value	Cluster Size	<i>p</i> _{corrected} value
		x	y	z				
MAIN EFFECTS								
<i>(a) Dance: Robotic > Natural</i>								
R inferior parietal lobule	40	63	-25	28	IPL	8.85	1898	<0.0001
R inf. parietal lobule	40	36	-37	49	IPL	8.35		
R postcentral gyrus	2	39	-31	43	S2	8.25		
L inferior parietal lobule	40	-51	-28	40	IPL	7.25	511	<0.0001
L ant. intraparietal sul.	40	-42	-37	49	aIPS	7.24		
L postcentral gyrus	2	-63	-19	16	S2	5.40		
L occipitotemporal cortex	37	-42	-70	-11	LOC	7.14	906	<0.0001
L mid. occipital gyrus	37	-51	-55	-5	MOG	6.74		
L fusiform gyrus	37	-30	-52	-20		6.71		
R inferior frontal gyrus	6/44	54	8	22	PM_v	4.82	84	0.035
R inferior frontal gyrus	44	42	2	13	IFG	3.63		
R precentral gyrus	6	27	-10	40	FEF	4.69	82	0.035
R middle frontal gyrus	6	27	5	61	MFG	4.55		
L precentral gyrus	6	-24	-4	49	FEF	4.49	62	0.058
L precentral gyrus	6	-30	2	58	PM _d	3.86		
R sup. temporal gyrus	22	42	-10	-5	STG	4.23	28	0.198
<i>(b) Dance: Natural > Robotic</i>								
R Lingual Gyrus	18	3	-82	-5	V2	8.45	739	<0.0001
R mid. occipital gyrus	18	24	-94	7	V2	8.36		
L occipital pole	17	-21	-94	13		7.35		
L anterior hippocampus	NA	-18	-7	-11		4.42	11	0.763
<i>(c) Face: Mask > Exposed</i>								
L inferior occipital gyrus	19	-27	-85	-8		6.63	78	0.142
L fusiform gyrus	37	-30	-70	-14		5.17		
R fusiform gyrus	37	33	-64	-11		4.66	16	0.626
INTERACTION								
<i>(d) Natural Dance (Face) + Robotic Dance (Mask) > Natural Dance (Mask) + Robotic Dance (Mask)</i>								
R lingual gyrus	30	12	-52	-2		5.87	131	0.065
R posterior cingulate	31	12	-58	16		5.28		
R parahippocampal gyrus	35	21	-40	-14		5.11	27	0.476
L lingual gyrus	19	-12	-49	-2		5.00	63	0.171
L parahippocampal gyrus	NA	-21	-37	-5		3.88		
R angular gyrus	39	42	-67	34	IPS	4.33	12	0.722
R inferior parietal lobule	39	48	-52	22	IPL	4.14	32	0.412

Locations in MNI coordinates and labels of peaks of relative activation from conditions of interest in Experiment 1. Regions more responsive to robotic than human-like dance are listed under (a), regions more responsive to human-like than robotic dance are listed under (b), and the main effect for mask > face is listed under (c). No suprathreshold activations emerged for the main effect of *exposed face > mask*. Only one direction of the interaction analysis reached significance, and is listed under (d). Results were calculated at $p_{\text{uncorrected}} < 0.001$, $k = 10$ voxels. Up to 3 local maxima are listed when a cluster has multiple peaks

more than 8mm apart. Entries in bold denote activations significant at the *FDR* cluster-corrected level of $p < 0.05$. Only regions that reached cluster-corrected significance are illustrated in the figures in the main text. Abbreviations for brain regions: IPL = inferior parietal lobule; S2 = secondary somatosensory cortex; aIPS = anterior intraparietal sulcus; LOC = lateral occipital complex; MOG = middle occipital gyrus; PMv = ventral premotor cortex; IFG = inferior frontal gyrus; FEF = frontal eye fields; MFG = middle frontal gyrus; PMd = dorsal premotor cortex; STG = superior temporal gyrus; V2 = visual area V2/prestriate cortex.

TABLE 2

Anatomical Region	BA	MNI			Putative Funct. Name	<i>t</i> value	Cluster Size	<i>p</i> _{corrected} value
		x	y	z				
MAIN EFFECTS								
<i>(a) Dance: Robotic > Natural</i>								
R inferior parietal lobule	40	63	-25	40	IPL	10.91	1687	<0.0001
R occipitoparietal cortex	19	42	-79	22		8.00		
R precentral sulcus	6	54	5	31	PM_v	7.33		
L occipitotemporal cortex	37	-36	-55	-8		9.80	384	<0.0001
L middle occipital gyrus	37	-45	-64	-14		8.96		
L fusiform gyrus	37	-22	-49	-17		6.30		
L inferior parietal lobule	40	-60	-31	34	IPL	9.15	399	<0.0001
L intraparietal sulcus	2	-45	-37	58	IPS	4.35		
R parahippocampal gyrus	37	33	-43	-11		7.88	452	<0.0001
R middle temporal gyrus	37	48	-49	-5		7.66		
R parahipp. gyrus	20	33	-25	-26		7.03		
L occipitoparietal cortex	19	-39	-82	19		6.00	47	0.090
L superior temporal gyrus	42	-33	-31	7	STG	5.48	240	<0.0001
L insula	NA	-36	-16	-5		4.79		
L sup. temporal gyrus	22	-45	-22	4	STG	4.47		
L inferior frontal gyrus	6/44	-51	-1	22	PM _v	5.12	37	0.119
R middle cingulate gyrus	24	6	-4	40	CMA/SMA	4.08	12	0.386
<i>(b) Dance: Natural > Robotic</i>								
R occipital pole	18	21	-97	1	V1/V2	9.85	880	<0.0001
L middle occipital gyrus	18	-27	-97	4	MOG	9.50		
R mid. occipital gyrus	18	24	-91	-8	MOG	8.16		
R lateral occipital cortex	19	48	-67	4	V5/hMT+	5.72	29	0.402
R superior frontal gyrus	6	6	8	67	SMA	4.38	29	0.402
L superior frontal gyrus	6	-6	-4	64	SMA	3.89		
<i>(c) Form: Lego > Human</i>								
R fusiform gyrus	19	30	-61	-17		10.87	2282	<0.0001
R lat. occipital cortex	19	36	-85	22	LOC	9.87		
R mid. occipital gyrus	19	27	-79	-8	LOC/V4_v	8.92		
L inferior occipital gyrus	19	-21	-94	-17	IOG	9.97	1839	<0.0001
L fusiform gyrus	37	-27	-64	-14		6.67		
L inf. occipital gyrus	19	-33	-94	-11	LOC	7.73		
L inferior parietal lobule	40	-54	-25	40	IPL	5.11	128	0.009
L intraparietal sulcus	2	-48	-40	55	IPS	4.52		
L ant. intraparietal sul.	2	-48	-34	46	aIPS	4.27		
R superior frontal gyrus	6	27	17	58	PM _d	4.86	53	0.080
R precentral gyrus	6	33	-10	49	FEF	4.32	53	0.080
R middle frontal gyrus	6	33	-7	58		4.20		
L superior frontal gyrus	6	-33	-1	58	PM _d	4.13	45	0.094
L precentral gyrus	6	-27	5	40		3.93		
L middle frontal gyrus	6	-51	2	40		3.97	15	0.376
INTERACTIONS								
<i>(d) (ND w/ HF) + (RD w/ LF) > (ND w/ LF) + (RD w/ HF)</i>								
L temporoparietal cortex	40	-51	-40	25	TPJ/IPL	5.47	35	0.566
<i>(d) (ND w/ LF) + (RD w/ HF) > (ND w/ HF) + (RD w/ LF)</i>								
R middle occipital gyrus	18	21	-97	-5		7.40	79	0.026
L inf. occipital gyrus	19	-21	-97	-17		5.66	45	0.051

Locations in MNI coordinates and labels of peaks of relative activation from conditions of interest in Experiment 2. Regions more responsive to robotic than human-like dance are listed under (a), regions more responsive to human-like than robotic dance are listed under (b), and regions more responsive to the lego form than the human form are listed under (c). No suprathreshold activations emerged for the main effect of *human form > lego form*. Both directions of the interaction are listed under (d) and (e). Results were calculated at $p_{\text{uncorrected}} < 0.001$, $k = 10$ voxels. Up to 3 local maxima are listed when a cluster has multiple peaks more than 8mm apart. Entries in bold denote activations significant at the *FDR* cluster-corrected level of $p < 0.05$. Only regions that reached cluster-corrected significance are illustrated in the figures in the main text. Abbreviations for interactions: ND = natural dance; RD = robotic dance; LF = Lego form; HF = human form. Abbreviations for brain regions: IPL = inferior parietal lobule; aIPS = anterior intraparietal sulcus; LOC = lateral occipital complex; MOG = middle occipital gyrus; PMv = ventral premotor cortex; IFG = inferior frontal gyrus; FEF = frontal eye fields; PMd = dorsal premotor cortex; STG = superior temporal gyrus; V2 = visual area V2/prestriate cortex; CMA = cingulate motor area; SMA = supplemental motor area; V1 = primary visual cortex; V5/MT = extrastriate visual cortex/middle temporal; IOG = inferior occipital gyrus; TPJ = temporoparietal junction.

Supplementary Figure 1.

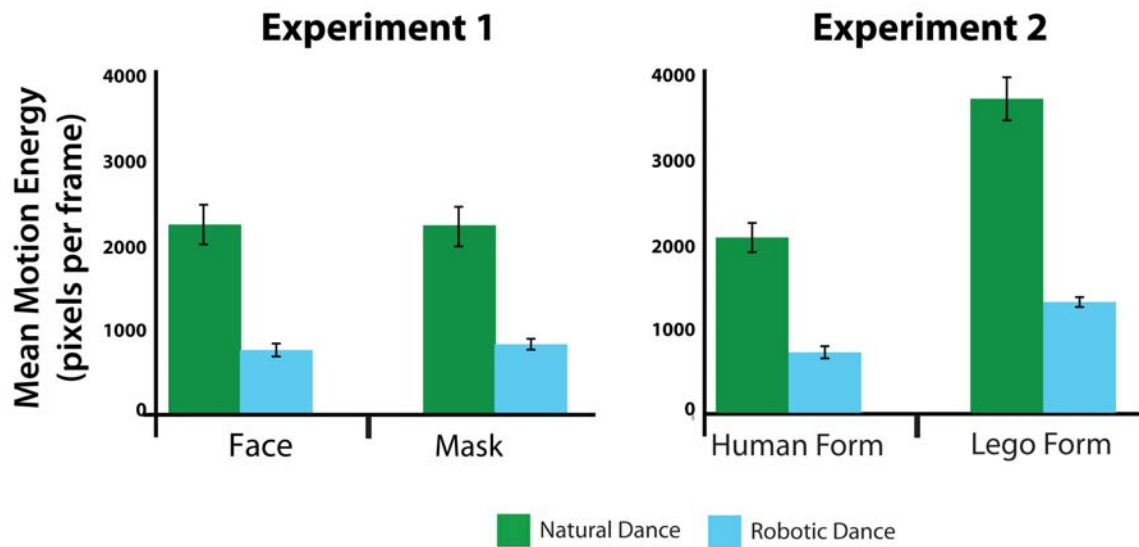


Figure S1. Results from motion energy quantification algorithm for all categories of stimuli, for Experiment 1 and Experiment 2. In Experiment 1, more motion energy emerged for natural dance compared to robotic dance ($F_{1,8} = 71.59, p < 0.001$). No main effect of whether the dancer was wearing a mask or had an exposed face was observed, nor was there a significant interaction between movement style and mask presence. In Experiment 2, more motion energy was recorded for videos featuring natural dance compared to robotic dance ($F_{1,7} = 106.03, p < 0.001$) and the Lego form compared to the human form ($F_{1,7} = 162.24, p < 0.001$). A significant interaction was also observed between both factors in Experiment 2, ($F_{1,7} = 45.72, p < 0.001$), signifying that the differences between motion energy for the natural and robotic dance for the Lego form were more pronounced than when the human was performing. As is explained in the main text, this is greater motion energy for the Lego form and what is likely driving the interaction is the fact that the Lego figure's limbs are proportionally longer than the human's (and the Lego figure's head also has several elongated spikes extending off the back), thus causing more pixels to be displaced between frames.

Supplementary Figure 2.

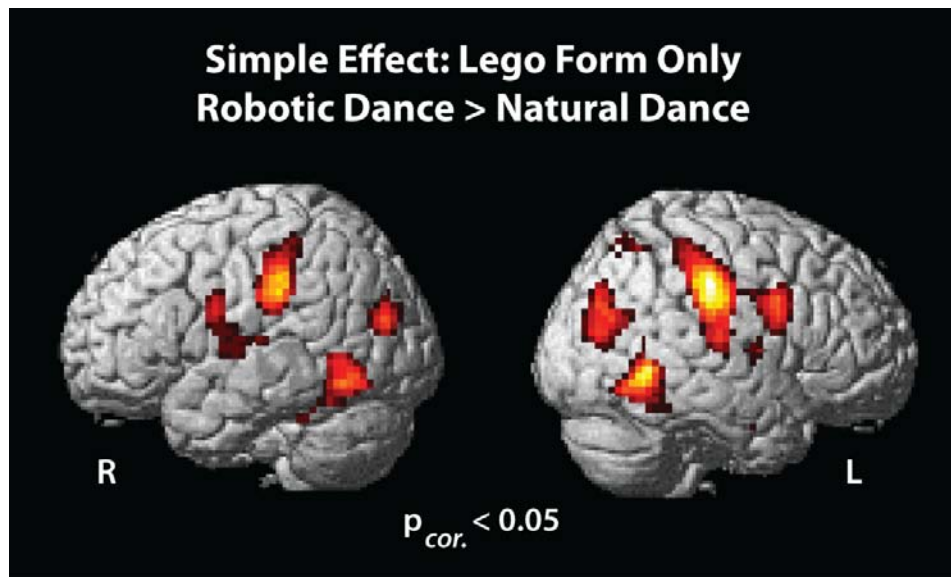


Figure S2. Simple effect of dance style, for Lego form only (Experiment 2). When just the two cells that featured the Lego form from Experiment 2 are compared, the same pattern as reported in the main effect (collapsed across human and Lego bodies) is replicated. Robust, bilateral activation within AON regions emerges when watching a Lego form dance robotically, compared to dancing like a human naturally would ($p_{corrected} < 0.05$).

REFERENCES

- Adolphs R. (2010): Conceptual challenges and directions for social neuroscience. *Neuron* 65(6):752-67.
- Aglioti SM, Cesari P, Romani M, Urgesi C. (2008): Action anticipation and motor resonance in elite basketball players. *Nat Neurosci*.
- Alaerts K, Swinnen SP, Wenderoth N. (2010): Observing how others lift light or heavy objects: which visual cues mediate the encoding of muscular force in the primary motor cortex? *Neuropsychologia* 48(7):2082-90.
- Avenanti A, Sirigu A, Aglioti SM. (2010): Racial bias reduces empathic sensorimotor resonance with other-race pain. *Curr Biol* 20(11):1018-22.
- Beer AL, Watanabe T, Ni R, Sasaki Y, Andersen GJ. (2009): 3D surface perception from motion involves a temporal-parietal network. *Eur J Neurosci* 30(4):703-13.
- Bobick AF. (1997): Movement, activity and action: the role of knowledge in the perception of motion. *Philos Trans R Soc Lond B Biol Sci* 352(1358):1257-65.
- Brass M, Bekkering H, Prinz W. (2001): Movement observation affects movement execution in a simple response task. *Acta Psychol (Amst)* 106(1-2):3-22.
- Brass M, Schmitt RM, Spengler S, Gergely G. (2007): Investigating action understanding: inferential processes versus action simulation. *Curr Biol* 17(24):2117-21.
- Buccino G, Lui F, Canessa N, Patteri I, Lagravinese G, Benuzzi F, Porro CA, Rizzolatti G. (2004): Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. *J Cogn Neurosci* 16(1):114-26.
- Calvo-Merino B, Glaser DE, Grezes J, Passingham RE, Haggard P. (2005): Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb Cortex* 15(8):1243-9.
- Chaminade T, Franklin DW, Oztop E, Cheng G. (2005): Motor interference between humans and humanoid robots: Effect of biological and artificial motion. *Proceedings of the 2005 4th IEEE International Conference on Development and Learning*:96-101.
- Chaminade T, Zecca M, Blakemore SJ, Takanishi A, Frith CD, Micera S, Dario P, Rizzolatti G, Gallese V, Umiltà MA. (2010): Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS One* 5(7):e11577.
- Chong TT, Williams MA, Cunnington R, Mattingley JB. (2008): Selective attention modulates inferior frontal gyrus activity during action observation. *Neuroimage* 40(1):298-307.
- Costantini M, Galati G, Ferretti A, Caulo M, Tartaro A, Romani GL, Aglioti SM. (2005): Neural systems underlying observation of humanly impossible movements: an FMRI study. *Cereb Cortex* 15(11):1761-7.
- Cross ES, Hamilton AF, Grafton ST. (2006): Building a motor simulation de novo: Observation of dance by dancers. *Neuroimage* 31(3):1257-67.
- Cross ES, Hamilton AF, Kraemer DJ, Kelley WM, Grafton ST. (2009a): Dissociable substrates for body motion and physical experience in the human action observation network. *Eur J Neurosci* 30(7):1383-92.

- Cross ES, Kraemer DJ, Hamilton AF, Kelley WM, Grafton ST. (2009b): Sensitivity of the action observation network to physical and observational learning. *Cereb Cortex* 19(2):315-26.
- Dapretto M, Davies MS, Pfeifer JH, Scott AA, Sigman M, Bookheimer SY, Iacoboni M. (2006): Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat Neurosci* 9(1):28-30.
- de Lange FP, Spronk M, Willems RM, Toni I, Bekkering H. (2008): Complementary systems for understanding action intentions. *Curr Biol* 18(6):454-7.
- Decety J, Sommerville JA. (2003): Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn Sci* 7(12):527-33.
- Engel A, Burke M, Fiehler K, Bien S, Rösler F. (2008): How moving objects become animated: The human mirror neuron system assimilates non-biological movement patterns. *Social Neuroscience* 3(3-4):368-387.
- Felician O, Anton JL, Nazarian B, Roth M, Roll JP, Romaiquere P. (2009): Where is your shoulder? Neural correlates of localizing others' body parts. *Neuropsychologia* 47(8-9):1909-16.
- Gallese V. (2007): Embodied simulation: from mirror neuron systems to interpersonal relations. *Novartis Found Symp* 278:3-12; discussion 12-9, 89-96, 216-21.
- Gallese V, Goldman A. (1998): Mirror neurons and the simulation theory of mindreading. *Trends Cogn Sci* 2:493-501.
- Gallese V, Keysers C, Rizzolatti G. (2004): A unifying view of the basis of social cognition. *Trends Cogn Sci* 8(9):396-403.
- Gazzola V, Keysers C. (2009): The Observation and Execution of Actions Share Motor and Somatosensory Voxels in all Tested Subjects: Single-Subject Analyses of Unsmoothed fMRI Data. *Cereb Cortex* 19(6):1239-55.
- 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):1674-84.
- Gowen E, Stanley J, Miall RC. (2008): Movement interference in autism-spectrum disorder. *Neuropsychologia* 46(4):1060-8.
- Grèzes J, Decety J. (2001): Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum Brain Mapp* 12(1):1-19.
- Hamilton AF, Grafton ST. (2006): Goal representation in human anterior intraparietal sulcus. *J Neurosci* 26(4):1133-7.
- Jackson PL, Meltzoff AN, Decety J. (2006): Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31(1):429-39.
- James W. (1890). *Principles of psychology*. New York: Holt.
- Jastorff J, Clavagnier S, Gergely G, Orban GA. (2011): Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cereb Cortex* 21(2):318-29.
- Kessler K, Biermann-Ruben K, Jonas M, Siebner HR, Baumer T, Munchau A, Schnitzler A. (2006): Investigating the human mirror neuron system by means of cortical synchronization during the imitation of biological movements. *Neuroimage* 33(1):227-38.

- Kilner J, Hamilton AF, Blakemore SJ. (2007a): Interference effect of observed human movement on action is due to velocity profile of biological motion. *Soc Neurosci* 2(3-4):158-66.
- Kilner JM, Friston KJ, Frith CD. (2007b): The mirror-neuron system: a Bayesian perspective. *Neuroreport* 18(6):619-23.
- Kilner JM, Friston KJ, Frith CD. (2007c): Predictive coding: an account of the mirror neuron system. *Cogn Process* 8(3):159-166.
- Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD. (2009): Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci* 29(32):10153-9.
- Kilner JM, Paulignan Y, Blakemore SJ. (2003): An interference effect of observed biological movement on action. *Curr Biol* 13(6):522-5.
- Liepelt R, Brass M. (2010): Top-down modulation of motor priming by belief about animacy. *Exp Psychol* 57(3):221-7.
- Liepelt R, Prinz W, Brass M. (2010): When do we simulate non-human agents? Dissociating communicative and non-communicative actions. *Cognition* 115(3):426-34.
- Liepelt R, Ullsperger M, Obst K, Spengler S, von Cramon DY, Brass M. (2009): Contextual movement constraints of others modulate motor preparation in the observer. *Neuropsychologia* 47(1):268-75.
- Liepelt R, Von Cramon DY, Brass M. (2008): How do we infer others' goals from non-stereotypic actions? The outcome of context-sensitive inferential processing in right inferior parietal and posterior temporal cortex. *Neuroimage*.
- Liew SL, Han S, Aziz-Zadeh L. (2010): Familiarity modulates mirror neuron and mentalizing regions during intention understanding. *Hum Brain Mapp*.
- Marsh LE, Hamilton AF. (2011): Dissociation of mirroring and mentalising systems in autism. *Neuroimage*.
- Meltzoff AN. (2007): 'Like me': a foundation for social cognition. *Dev Sci* 10(1):126-34.
- Miura N, Sugiura M, Takahashi M, Sassa Y, Miyamoto A, Sato S, Horie K, Nakamura K, Kawashima R. (2010): Effect of motion smoothness on brain activity while observing a dance: An fMRI study using a humanoid robot. *Soc Neurosci* 5(1):40-58.
- Neal A, Kilner JM. (2010): What is simulated in the action observation network when we observe actions? *Eur J Neurosci*.
- Oberman LM, Hubbard EM, McCleery JP, Altschuler EL, Ramachandran VS, Pineda JA. (2005): EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Brain Res Cogn Brain Res* 24(2):190-8.
- Obhi SS, Hogeveen J. (2010): Incidental action observation modulates muscle activity. *Exp Brain Res* 203(2):427-35.
- Oldfield RC. (1971): The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97-113.
- Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE. (2010): Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *J Neurophysiol* 104(2): 1077-89.
- Peelen MV, Downing PE. (2007): The neural basis of visual body perception. *Nat Rev Neurosci* 8(8):636-48.

- Press C. (2011): Action Observation and Robotic Agents: Learning and Anthropomorphism. *Neurosci Biobehav Rev*.
- Prinz W. 1990. A common coding approach to perception and action. In: Neumann O, Prinz W, editors. Relationships between perception and action: Current approaches. Berlin: Springer-Verlag. p 167-201.
- Ramsey R, Hamilton AFdC. (2010): Triangles have goals too: understanding action representation in left aIPS. *Neuropsychologia* 48(9):2773-6.
- Rizzolatti G, Craighero L. (2004): The mirror-neuron system. *Annu Rev Neurosci* 27:169-92.
- Schippers MB, Roebroek A, Renken R, Nanetti L, Keysers C. (2010): Mapping the information flow from one brain to another during gestural communication. *Proc Natl Acad Sci U S A* 107(20):9388-93.
- Schubotz RI. (2007): Prediction of external events with our motor system: towards a new framework. *Trends Cogn Sci* 11(5):211-8.
- Shimada S. (2010): Deactivation in the sensorimotor area during observation of a human agent performing robotic actions. *Brain Cogn* 72(3):394-9.
- Spunt RP, Satpute AB, Lieberman MD. (2011): Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. *J Cogn Neurosci* 23(1):63-74.
- Stanley J, Gowen E, Miall RC. (2010): How instructions modify perception: an fMRI study investigating brain areas involved in attributing human agency. *Neuroimage* 52(1):389-400.
- Tai YF, Scherfler C, Brooks DJ, Sawamoto N, Castiello U. (2004): The human premotor cortex is 'mirror' only for biological actions. *Curr Biol* 14(2):117-20.
- Van Oostende S, Sunaert S, Van Hecke P, Marchal G, Orban GA. (1997): The kinetic occipital (KO) region in man: an fMRI study. *Cereb Cortex* 7(7):690-701.
- Wheatley T, Milleville SC, Martin A. (2007): Understanding animate agents: distinct roles for the social network and mirror system. *Psychol Sci* 18(6):469-74.
- Zacks JM, Ollinger JM, Sheridan MA, Tversky B. (2002): A parametric study of mental spatial transformations of bodies. *Neuroimage* 16(4):857-72.
- Zentgraf K, Stark R, Reiser M, Kunzell S, Schienle A, Kirsch P, Walter B, Vaitl D, Munzert J. (2005): Differential activation of pre-SMA and SMA proper during action observation: effects of instructions. *Neuroimage* 26(3):662-72.