Acting together in and beyond the mirror neuron system

Idil Kokal, Valeria Gazzola, Christian Keysers *

BCN NeuroImaging Center, University of Groningen, The Netherlands
Department of Neuroscience, University Medical Center Groningen, The Netherlands

ARTICLE INFO

ABSTRACT

Moving a set dinner table often takes two people, and doing so without spilling the glasses requires the close coordination of the two agents' actions. It has been argued that the mirror neuron system may be the key neural locus of such coordination. Instead, here we show that such coordination recruits two separable sets of areas: one that could translate between motor and visual codes and one that could integrate these information to achieve common goals. The former includes regions of the putative mirror neuron system, the latter, regions of the prefrontal, posterior parietal and temporal lobe adjacent to the putative mirror neuron system. Both networks were more active while participants cooperated with a human agent, responding to their actions, compared to a computer that did not, evidencing their social dimension. This finding shows that although the putative mirror neuron system can play a critical role in joint actions by translating both agents' actions into a common code, the flexible remapping of our own actions with those of others required during joint actions seems to be performed outside of the putative mirror neuron system.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Joint actions are "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz et al., 2006a).

As very few studies have investigated brain activity during joint actions (Newman-Norlund et al., 2007a, 2008; Sebanz et al., 2006b, 2007) the brain mechanisms supporting joint actions are still unknown. In comparison, there are more studies investigating brain activity while participants executed an action or/and observed an action. In the monkey, single cell recordings showed that some neurons, called mirror neurons, were active both during the observation and execution of similar actions (Fijii et al., 2007; Fogassi et al., 2005; Gallese et al., 1996; Keysers et al., 2003; Kohler et al., 2002; Umilta et al., 2001). In humans, voxels in similar locations have been found to be active during action observation and execution and form what should be called the *putative* mirror neuron system (Gazzola et al., 2006, 2007a,b; Gazzola and Keysers, 2009; Grafton et al., 1996; Hamilton et al., 2007; Iacoboni and Dapretto, 2006; Keysers and Gazzola, 2006; Rizzolatti and Craighero, 2004; Rizzolatti et al., 1996). The term 'putative' here underlines the fact that if a voxel in an fMRI experiment shows an increase in BOLD both during action observation and execution, this suggests that it could contain mirror neurons, but it could also contain distinct but interdigitated populations of neurons involved in action observation only and execution only, commending

caution in interpretation (Gazzola and Keysers, 2009). The monkey mirror neuron system and its putative human equivalent have been implicated in many aspects of social interactions, including imitation (Iacoboni et al., 1999), empathy and simulation (Fijii et al., 2007; Fogassi et al., 2005; Gallese et al., 1996; Gazzola et al., 2006, 2007a,b; Gazzola and Keysers, 2009; Keysers et al., 2003; Kohler et al., 2002; Umilta et al., 2001), mind-reading (Gallese, 2003; Gallese and Goldman, 1998) and language (Rizzolatti and Arbib, 1998).

Recently, the putative mirror neuron system was proposed to play a central role in joint actions because of the close link between perception and action provided by these brain regions (Knoblich and Jordan, 2002; Newman-Norlund et al., 2007a, 2008). According to this proposal, actors use simulation to predict the intentions and consequences of the actions of their co-actor. This would help the actor adjust his own action plans to the predicted actions of co-actor in order to successfully achieve a joint goal. Going one step further, Newman-Norlund et al. (2008), in a virtual lifting task, found that the BOLD signal in the right inferior frontal gyrus (IFG) was larger while participants balanced a ball together with another agent (joint action) compared to when they balanced the ball alone. The authors suggest that this finding indicates a direct role of the putative mirror neuron system in the integration of observed and executed actions during joint actions. However much of the IFG does not have mirror properties, and given that the authors have not mapped the putative mirror neuron system of their participants, the IFG but not the putative mirror neuron system may be responsible for this effect (Thioux et al., 2008).

Here we propose that a typical joint action requires at least three, more or less intertwined but conceptually separable processes: observing the actions of others (*observation*), executing ones own

---

* Corresponding author. University Medical Center Groningen, University of Groningen, Department of Neuroscience, BCN NeuroImaging Center, A. Deusinglaan 2, 9713 AW Groningen, The Netherlands. Fax: +31 50 3638875.
  E-mail address: c.m.keysers@rug.nl (C. Keysers).

actions (*execution*) and integrating *obs* and *exe* to tune one's own actions to those of others (*integration*).

In contrast to the interpretation of Newman-Norlund et al. (2008), we hypothesize that the putative mirror neuron system does not participate in the integration component of joint actions. This is because the neurons in the monkey's premotor cortex that have been described to respond to the sight of other people's actions show a fixed relationship between effective observed and executed actions (Gallese et al., 1996). For both strictly and broadly congruent mirror neurons, this relationship is one of correspondence, with 'correspondence' meaning that the actions have the same goal (broadly congruent, 60.9%) or the same goals and means (strictly congruent, 31.5%). For the minority (7.6%) of 'non-congruent' or 'logically related' visuo-motor neurons this relationship is different, and can include complementarity (e.g. execution of grasping and observation of placing), but again, this relationship is fixed over trials. By fixed relationship, we do not preclude the fact that extensive training can change this relationship (Catmur et al., 2007), but that it is not known to change in seconds based on task demands. The putative mirror neuron system could therefore promote joint actions by constantly linking the observation of actions to the motor programs for similar or complementary actions (Iacoboni et al., 1999; Rizzolatti and Craighero, 2004), but the integration needed in joint actions has to be more flexible: during joint actions, unlike imitation, the task determines the nature of this integration, which can vary from doing the same (e.g. lifting a table together) to the opposite of a partner (e.g. moving a table with one partner moving forwards and one backwards) in seconds. This rapid task-dependent redefinition of the visuo-motor integration goes beyond the known properties of mirror neurons and logically related visuo-motor neurons and is likely to recruit separate brain regions.

Testing this hypothesis therefore necessitates (i) a joint action task with trials requiring doing a similar action and trials requiring doing the opposite of a partner to achieve a common goal and (ii) observation (*obs*) and execution (*exe*) control tasks to map the putative mirror neuron system in the same participants. In Experiment I we therefore introduced a novel joint action paradigm (Fig. 1 and Methods and materials) that encompasses conditions in which participants only observe or only execute solo actions as well as joint action conditions in which they additionally have to integrate observation and execution by executing an action similar or opposite to the one observed. We identified regions involved in *integration* during joint actions by requiring that activity in joint actions exceeds that during solo observation plus execution (if *integration* > 0 then *joint action* (= *obs* + *exe* + *integration*) > *obs* + *exe*). This requirement is similar to the criterion of superadditivity used to identify regions involved in multisensory integration (Beauchamp, 2005), and we will therefore abbreviate regions showing evidence for superadditive integration during joint actions as 'sJA'. It also resembles the definition of imitation selective areas introduced by Iacoboni et al. (1999): *imitation* > *exe* + *obs*. The location of this network involved in *integration* can then be compared with that of the network of the putative mirror neuron system defined as voxels active during observation and execution. Finding none or minimal overlap between the networks would support our hypothesis that the putative mirror neuron system is not the primary locus of the *integration* process in joint actions. In contrast, finding that the sJA network falls within the putative mirror neuron system, particularly in the IFG, would support Newman-Norlund et al.'s (2008) interpretation that the *integration* is computed within the putative mirror neuron system network.

Furthermore, joint actions in the strict sense require mutual coordination between two agents. While shooting clay pigeons for instance, we need to adapt our own actions to the movements of an object in the outside world, which is an example of one-way coordination of an agent to an event in the outside world. This, however, does not qualify as a joint action because clay pigeons do not react to our movements. In contrast, lifting a table together does

qualify, because the lifters mutually coordinate their movements to one another's. To examine whether brain regions identified in Experiment I are sensitive to this distinction, in Experiment II we scanned half of the participants a second time, while playing the same cooperation game (a) with another person that adapts her movement to those of the player (mutual coordination, true joint actions) or (b) with a computer that does not (one-way coordination).

Materials and methods

Experiment I

Participants

18 healthy volunteers (all right-handed; 10 female and 8 male; mean age 23.7 years ranging 20–45 years) with normal or corrected to normal vision and without a history of neurological, major medical, or psychiatric disorders. The experiment was approved by the Medical Ethical Commission of the University Medical Center Groningen, the Netherlands. Participants gave informed consent and were paid for their participation.
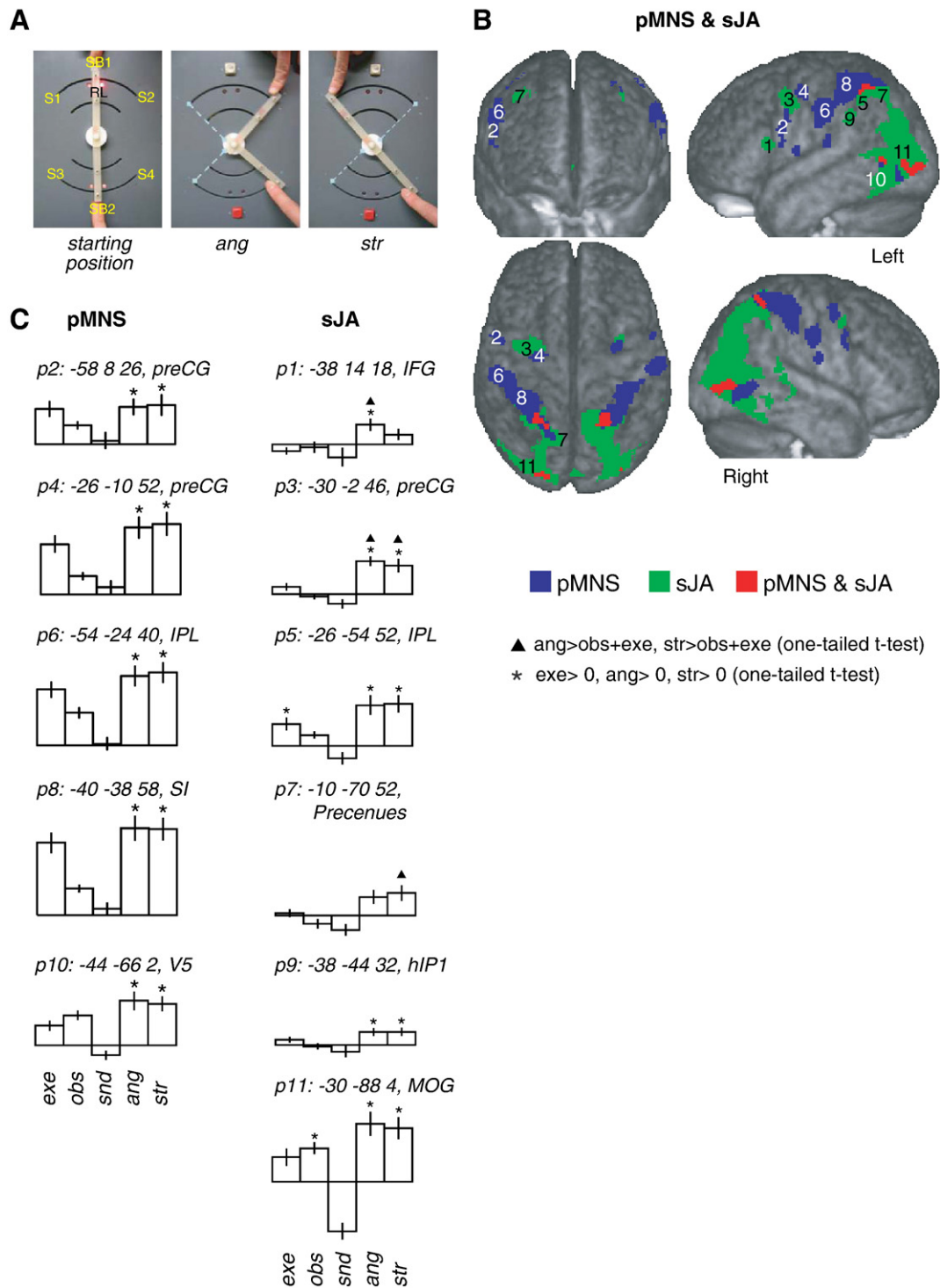
Response box

We used a custom-made MRI-compatible response box. The box was placed on the lower abdomen of the participant who was lying on the scanner bed (Fig. S1A). Using the mirror of the head coil participants were able to see the entire response box, their own fingers as well as the fingers of the experimenter who was standing next to the scanner bed. The participant and the experimenter wore MRI-compatible headphones for auditory instructions. The response box had 2 individual arms made of fiberglass sticks resembling the hands of an analog clock with an hour hand and a minute hand of equal length (11 cm, Fig. S1B). At rest, the experimenter's stick pointed up (12:00 of an analog clock) and the participant's down (06:00). Four sensors ('S1'–'S4' in Fig. S1B) placed at 2, 4, 8 and 10 o'clock, measured the time point at which the experimenter or participant reached the target position with their stick by pushing the stick with their index finger. The spring-loaded sticks returned to the initial positions (12 and 6 o'clock) when released at the end of the trial. Two starting buttons ('SB1' and 'SB2' in Fig. S1B) at locations 12 and 6 o'clock served as 'home base' for the experimenter and participant respectively (Fig. S1B), and red LEDs ('RL' in Fig. S1B) were turned on for as long as the experimenter or participant pressed her or his starting button.

Procedure

In the fMRI session, the experimenter and the participant (both right-handed) played a cooperation game or performed one of three non-cooperative control conditions while the participant was being scanned.

The task of the participant in the joint action conditions was to cooperate with the experimenter to shape the two sticks of the box in either an angle or a straight line (see Figs. S1C, D and video S1). At the beginning of each trial, both players had their index finger on their starting button (SB1 and SB2 in Fig. S1B). Before the actual movement, the experimenter received auditory instructions indicating where (left or right) and when to move her stick. The participant was unaware of these instructions, and received auditory instructions only indicating whether to create an angle (*ang*) or a straight line (*str*, Fig. S2). The experimenter, present in the scanner room, initiated the cooperation by moving the top arm of the clock to the left or right. The participant had to react by starting to move the lower arm of the clock in the direction suitable to achieve the target shape (Video S1). More specifically, the participant had to release his start button after the experimenter had left hers, which occurred between 1 and 2 s (random interval) after the participant had received the angle or straight

Fig. 1. Experimental design and results. (A) Left: photograph of the response box together with the fingers of the experimenter (top) and participant (bottom); middle: correct configuration for an angle trial, dotted lines showing alternative configuration; right: same for a straight trial. (B) Rendering of average brain of participants with pMNS (putative mirror neuron system) (blue, *exe* > 0 and *obs* > 0, both *p* < 0.001), sJA (superadditive voxels in joint actions) (green, a global null conjunction of *str* > *exe* + *obs* + *snd* and *ang* > *exe* + *obs* + *snd* inclusively masked with (*str* > *exe* and *str* > *obs* and *str* > *snd*) or (*ang* > *exe* and *ang* > *obs* and *ang* > *snd*) see Materials and methods) and overlap between pMNS and sJA (red). Numbers refer to the location of peak parameter extraction in C. (C) Parameter estimates of the peak locations in sJA and putative mirror neuron system in the left hemisphere, with 'p1' referring to peak location 1 in panel b and the triplet of numbers indicating the MNI coordinates of the peak. Parameter estimates were compared against zero (one-tailed *t*-test) for *ang* and *str* conditions in putative mirror neuron system regions (left column) and for *exe*, *obs*, *ang* and *str* for sJA regions (right column), and stars denote significant results. Finally, in all peak locations the comparison *ang* − *str* was not significant (two-tailed comparison). All comparisons thresholded at *p* < 0.01 corrected for the number of peak locations in which the comparison was performed. All parameter estimates and the error bars (SEM) are shown on the same scale and can be directly compared (actual units irrelevant because arbitrary). Parameter estimates of right hemisphere are not shown because they are virtually identical.

instruction. Thus, s/he had to carefully watch the experimenter's actions to determine (a) when the experimenter started her movement and (b) which side the experimenter moved towards. This allowed the participant to determine when to start his/her own action and, in combination with the knowledge of the target shape (straight or angle), which side to move his/her stick towards. The experimenter and

participant then had to reach the target location virtually simultaneously (within 200 ms of each other) to jointly win the trial. This tight time constraint ensured that the participant and experimenter had to monitor and coordinate the velocity of their movements carefully and continuously throughout the trial, requiring both the spatial and temporal coordination that defines joint actions. It also makes our laboratory paradigm similar to real-world joint actions such as lifting a dinner table, where the velocity of actions have to be coordinated to avoid tipping over the objects. The experimenter varied her initial movement velocity from trial-to-trial and participants responded to these changes showing that they indeed adjusted their own actions continuously to those of the experimenter (see Video S1). Thereafter, both agents had to, and did, adjust their movements to the velocity of the other to meet the common goal of reaching the target location within 200 ms of each other, conveying a mutual feeling of cooperation.

After the end of a trial, both players had to place their index finger back onto their respective starting button. At the end of each run, the participant and experimenter were informed about how many points they have jointly earned to maintain motivation. At the end of the experiment, we debriefed the participants about the experiment. We asked the following questions: Did you perceive the trials as games? Did you try to cooperate with the experimenter? Was she cooperating with you? Which game did you find harder; angle or straight? At debriefing, all participants reported perceiving the task as a cooperation game, trying to be as successful as possible and feeling that the experimenter adjusted her movement to match theirs and vice versa. None of the participants made more than 3 errors out of the 72 game trials played per run. Unfortunately, reaction time data is not available because for 10/18 participants, one of the two target locations were hard to reach within the space constraints of the scanner, and they were instructed that the direction and timing of the movement were more critical than reaching the actual target location. For these participants, during scanning, the experimenter kept track of the number of direction errors (i.e. not going towards the experimenter's side in angle or going to the experimenter's side in straight trials), and verbally informed the participants of their performance. Inspection of the data from these 10/18 participants however did not suggest any systematic differences with those of the remaining 8 participants.

The experiment contained 6 conditions that were arranged in blocks of 8 trials lasting between 45 and 54 s (depending on the random intervals separating trials, Fig. S2).

1) *Angle (ang)*: 8 trials separated by 2.3 s all starting with the instruction 'angle' (450 ms with a 150 ms silence added at the end to match the length of 'straight' sound).
2) *Straight (str)*: as in *ang*, but all trials started with the instruction 'straight' (600 ms).
3) *Mixed (mix)*: 4 angle trials randomly intermixed with 4 straight trials. Blocks of types 1–3 involve joint actions and 1.75 s before each block a 130 ms tone (sine wave, 440 Hz) instructed participants that they would have to play the cooperation game. In each block, the experimenter moved her stick 4 times to the right and 4 times to the left, in random order.
4) *Sound (snd)*: participants heard the 'angle' and 'straight' instructions using the exact same timing as in a *mix* block. 1.75 s before the onset of a *snd* block, participants heard a verbal instruction 'eyes close' (900 ms) ordering them to close their eyes and indicating that the next block required them to listen to the auditory instructions without further actions. 1500 ms after the end of *snd* blocks, a voice stating 'eyes open' (900 ms) instructed participants to reopen their eyes.
5) *Observation (obs)*: participants only viewed the experimenter move her stick to the right or left using the exact same timing as in a joint action block. 1.75 s before the block, the verbal instruction 'look' (400 ms) instructed the participants only to observe the experimenter.

6) *Execution (exe)*: In conditions 1, 2, 3 and 5, a red light (RL in Fig. S1B) was turned on whenever the experimenter placed her finger on the start button (SB in Fig. S1B) and turned off whenever she left the SB to start her action. In the execution condition, the experimenter's RL was turned on and off with the same timing as in the conditions 1, 2, 3 and 5 without the experimenter being visible. The participant had to move his/her stick to the right or left whenever he/she saw the red light turn off on the box, ensuring that the timing of the participant's actions was the same as in the joint action blocks but not triggered by a biological action. The participant could choose which side to go to, but was instructed by the experimenter to avoid going to the same side constantly. A verbal instruction 'action' (400 ms) presented 1.75 s before the block indicated the nature of the block.

Blocks were separated by 14 ± 2 s random pauses (including the verbal instruction or sound indicating the type of block to follow). Each run lasted 720 s and contained 2 blocks of each of the 6 conditions and a feedback at the end. Five runs were acquired, for a total of 10 blocks of each condition. The order of the conditions was counterbalanced between runs and participants. Stimuli were programmed and presented using the Presentation software (Neurobehavioral systems, Davis, CA).

Participants were familiarized with all the conditions during a training session performed outside of the scanner on a separate day. This training was composed of three 720 s sessions identical to those used in the main experiment. In the third of these sessions, all participants' performance was perfect (i.e. not a single error in the 72 trials). None of the participants reported being confused about the conditions during the main experiment in the scanner. This training session also ensured that those participants who were unable to reach the sensors in the fMRI experiment were familiar with the time constraints of the game, which were accurately measured out of the scanner, but impossible to measure during scanning for these participants.

Data acquisition

Imaging was performed with a Philips Intera 3 T Quaser with a synergy SENSE head coil and maximum gradient strength of 30 mT/m. Head movements were minimized by using foam padding and never exceeded 3 mm in a run. We used a standard single shot EPI with TE = 28 ms, TA = 1.25 s, TR = 1.3 s, 28 axial slices of 4 mm thickness, without slice gap and a 3.5×3.5 mm in plane resolution acquired to cover the entire brain. The first 5 volumes of each functional run were discarded for the longitudinal magnetization to approach equilibrium. A T1 weighted structural scan was acquired with TR = 15.31 ms, TE = 3.6 ms, and flip angle = 8°.

Data preprocessing

Using SPM2 (www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 6.5 (Mathworks Inc., Sherborn, MA, USA), all EPI volumes were aligned to the first volume acquired for each participant and a mean EPI image was generated after realignment. Spatial normalization was performed by co-registering the structural volume to the mean EPI, segmenting the coregistered structural image, determining the normalization parameters required to warp the gray matter segment onto the gray matter MNI template, and applying these parameters to all EPI and structural volumes. The normalized EPI images were smoothed with a 6 mm FWHH isotropic Gaussian kernel.

General data analyses

Functional data were analyzed using a general linear model (GLM) separately for each participant and voxel using SPM2. Although the experiment was presented as a block design, we modeled the data in an event related fashion because examination of the signal time course within the blocks showed clearly visible peaks for each trial.

Single participant analyses

The GLM was performed using separate auditory predictors for the conditions *ang*, *str*, *mix* and *snd* to capture brain activity caused by hearing the words "angle" or "straight" and separate action predictor for the *ang*, *str*, *mix*, *obs* and *exe* conditions to capture brain activity triggered by executing and/or observing the finger movements. Each predictor was a boxcar function that reflected the trial-by-trial timing of the auditory and movement epoch of the condition (much as in Fig. S2, but the action predictor corresponded to the union of the experimenter's and participant's action time course). The boxcar functions were convolved with the haemodynamic response function, and fitted separately for each run to the data. In addition, the head motion and rotation along the three axes were entered as 6 covariates of no interest in the design matrix to single out motion artifacts although motion never exceeded 3 mm within a run. Given that little time separated the auditory instructions from the actions within a block (average = 1500 ms), the auditory and action predictors overlap in time (after convolution with the hrf), and the attribution of a brain activity to one rather than the other uncertain. Instead of analyzing the parameter estimates for the auditory and action predictor separately, we therefore combined them by summing the surface under the fitted auditory and action predictors. This was done simply by multiplying the parameter estimates (Beta) obtained from the GLM with the surface (S) under their respective predictor ($S = \mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}}$). Brain activity across conditions can then be compared using this surface. For instance whether for a particular voxel, the activity in the *str* condition exceeds that of the sum of the *snd*, *exe* and *obs* conditions, a contrast value $C = (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{str}} - (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{snd}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{obs}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{exe}}$ can be calculated and tested using the null hypothesis $C = 0$. Note that S_{auditory} and S_{action} are relatively constant across participants and conditions (angS_{auditory} average = 16, SEM = 0.0002 and angS_{action} average = 40.06, SEM = 0.86, strS_{auditory} average = 16, SEM = 0.0029 and strS_{action} average = 40.14, SEM = 0.88, in arbitrary units) because the timing of the conditions was relatively constant.

Population analyses

At the second level of analysis, to implement a random effect analysis, contrast estimates obtained separately for each participant were tested at the population level, using one-sample *t*-tests or analyses of variances (ANOVA) that test whether the average contrast differs from zero. Only results that are significant both at $p < 0.001$ uncorrected and $p < 0.05$ corrected using false discovery rate are reported as significant. Only clusters of at least 10 voxels are shown.

pMNS (putative mirror neuron system) definition

In particular, to determine voxels involved in the putative mirror neuron system, the surface under the curve in *obs* was compared against zero (*t*-test), and the same was done for *exe*, and only those voxels with significant results in both analyses at the second level were considered to belong to the putative mirror neuron system: $(\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{obs}} > 0$ & $(\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{exe}} > 0$, where & is logical, both at $p_{\mathrm{unc}} < 0.001$ and $p_{\mathrm{fdr}} < 0.05$ (here, using the surface under the curve or the parameter estimate alone is mathematically virtually equivalent because S_{action} was very similar across participants). This operational definition is far from perfect: a voxel can be involved in both execution and observation although the individual neurons within that voxel are not involved in both, which is why we refer to these voxels, not as 'mirror', but as *putative* mirror. This definition is however relatively well established in the neuroimaging literature (Chong et al., 2008; Gazzola et al., 2006, 2007a,b) and is the most direct translation of the original definition at the single cell level in monkeys (Gallese et al., 1996; Keysers et al., 2003; Kohler et al., 2002; Umilta et al., 2001). A similar definition is also used in domains of emotions (Jabbi et al., 2007; Singer et al., 2004, 2006; Wicker et al., 2003) and sensations (Blakemore et al., 2005; Keysers and Perrett, 2004; Keysers et al., 2004). Given that the main point of the present paper is that joint actions require more than the putative mirror neuron system alone, showing this while running the risk of overestimating the extent of the putative mirror neuron system (e.g. by including voxels that contain intermixed populations of responding to only the observation or only the execution of actions or responding to less general factors such as attention) actually strengthens the point.

sJA (superadditive voxels in joint actions) definition

To map regions showing activity that indicates their contribution in integrating observed and executed actions, two contrasts were calculated at the first level ($C_{\mathrm{ang}} = (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{ang}} - (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{snd}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{obs}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{exe}}$ and $C_{\mathrm{str}} = (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{str}} - (\mathrm{Beta}_{\mathrm{auditory}} \times S_{\mathrm{auditory}} + \mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{snd}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{obs}} - (\mathrm{Beta}_{\mathrm{action}} \times S_{\mathrm{action}})_{\mathrm{exe}}$). This definition included the *obs* and *exe* conditions as outlined in the introduction plus the *snd* condition to control for the effects of the auditory instructions given during the joint action tasks. Again, the logic of this definition is that what distinguishes joint actions from solo actions is that joint actions not only include the processes of observing actions and executing actions, but also additionally requires partners to *integrate* these two processes to achieve a goal. This integration is what the sJA maps try to capture. They do so by deducing that if *joint action = obs + exe + integration (+snd)*, and *integration > 0*, then *joint action − obs − exe − snd = integration* and has to be >0. Note that since the baseline condition between trials and blocks included the vision of the game box, the parameter estimates for the various conditions should capture deviations from this baseline, namely the sight of the experimenter's finger action in *obs*, and the motor control and observation of the participant's own finger movement in the *exe* condition. Subtracting both *obs* and *exe* from joint actions therefore does not represent a double subtraction of the visual input, but two separate subtractions of the unique visual components. The 18 C_{ang} and the 18 C_{str} contrasts (one per participant) were entered in a one-way ANOVA without constant, and the global null conjunction calculated to estimate the likelihood of the null hypothesis ($m(C_{\mathrm{ang}}) <= 0$ & $m(C_{\mathrm{str}}) <= 0$) that the voxel was not involved in either joint action (Friston et al., 2005). A global null hypothesis is appropriate here because unlike the definition of the putative mirror neuron system requiring that both *exe* and *obs* be above zero, for a voxel to be involved in joint actions it is sufficient that it be involved in one of the two joint actions.

Examining the parameter estimates in the resulting SPM showed however that sJA contained voxels in which the activity in *ang* or *str* was above *exe + obs + snd* without being above the activity in *exe*, *obs* and *snd* taken individually. This was the case for instance when *obs* and *snd* had negative parameter estimates, because *exe + obs + snd* was then less than *exe* alone. To prevent this unwanted effect, we additionally required that sJA voxels fall within an inclusive mask where ((*ang > exe* and *ang > obs* and *ang > snd*) or (*str > exe* and *str > obs* and *str > snd*)), with each individual contrast in that logical conjunction taken at $p < 0.05$ uncorrected.

sJA′ (alternative sJA) definition

To examine whether reductions in BOLD during *snd* only could have artificially inflated the number of sJA, we also examined which voxels satisfy a criterion excluding *snd* (i.e. the contrast *joint action − (exe + obs)* instead of *joint action − (exe + obs + snd)*, see Fig. S3A).

maxJA (joint action voxels calculated using a maximum requirement) definition

As an alternative to the sJA definition that rests on a criterion of superadditivity, we explored the impact of using what has been called a "maximum requirement" as well (Beauchamp, 2005). This

requirement typically states that multisensory response has to be larger than the maximum of the unisensory responses (Beauchamp, 2005). Adapted to our situation we therefore require that joint action related activity be more than the maximum activity in *observation* and *execution* of solo actions (*ang*>max (*exe,obs*) or *str*>max (*exe,obs*). In order to do so, for each participant we calculated two contrast maps using ImCalc in SPM: *ang* − max (*exe,obs*) and *str* − max(*exe,obs*). Second, we created a second level ANOVA that included these two contrast maps for each participant. Third, we performed a global null conjunction between the *ang* − max (*exe,obs*) and *str* − max (*exe, obs*). This map was then thresholded at *p*<0.001 uncorrected (which survived also to *p*<0.05 fdr correction). Results reflect those voxels in which either the *ang* or the *str* condition (or both) exceeded the maximum of *exe* and *obs*.

Analysis of the peak voxels

To compare the functional properties of the putative mirror neuron system and sJA regions, we selected the location of the peak *obs* activations in the main putative mirror neuron system clusters (dorsal and ventral premotor and parietal clusters and MTG) and identified locations of peak activity according to the global null conjunction in the sJA that were in anatomically similar locations (dorsal and ventral frontal and parietal, precuneus and high-level visual cortex). Peaks were extracted both from the left and right hemispheres in corresponding locations but the pattern in the right hemisphere was so similar to that in the left hemisphere, that only those of the left hemisphere are shown in Fig. 1C. We then extracted the mean signal time course in these peak voxels and analyzed this signal using a GLM with the same predictors used for the voxel-by-voxel analysis but using MarsBar (http://marsbar.sourceforge.net). The surface esti-mates of the conditions as defined above together with the MNI coordinates of the peaks were then plotted in Fig. 1C. Thereafter, a number of planned comparisons were performed at the second level based on the surface estimates of the 18 participants. In particular, for sJA peak voxels, we tested whether the activity in *exe*, *obs*, *ang*, and *str* exceeded zero (one-tailed *t*-test using a threshold of 0.01/6 to correct for the fact that the test was performed in 6 peak voxels) to check whether the ROIs behave as would be expected for the putative mirror neuron system and whether the joint action conditions show activity above baseline. Additionally we tested in sJA peak voxels whether the activity in *ang* and *str* exceed the sum of *exe* + *obs* (contrast *ang* > *exe* + *obs* and *str* > *exe* + *obs*, one-tailed *t*-test) to examine whether the voxels designated to joint action showed greater activation during joint actions than the sum of *exe* + *obs* independent of deactivations during *snd* condition (triangles in Fig. 1C). For putative mirror neuron system peak voxels, we tested whether the parameter estimates in the *ang* and *str* condition exceeded zero to check whether the putative mirror neuron system is involved in joint actions (the result was always significant) and whether the activity in the *ang* condition and the *str* conditions exceeded the sum of *exe*, *obs* and *snd* (contrast *ang* − *exe* − *obs* − *snd* and *str* − *exe* − *obs* − *snd*, results were never significant) using one-tailed *t*-test with a threshold of 0.01/5 to correct for 5 peak locations in all 4 tests. Finally, in all 11 peak voxels we examined the effect of rule switching by examining if the mixed condition exceed the unmixed joint actions conditions (contrast 2*mix*>*ang* + *str*, one-tailed *t*-test against zero, threshold of 0.01/11) and we compared *ang* and *str* conditions (*ang* − *str*, two-tailed *t*-test) (Fig. S4B).

Results were not Bonferroni corrected for the number of comparisons (8 for sJA and 6 for putative mirror neuron system) because they were planned a priori. We also extracted the mean activity from the entire clusters of subthreshold voxels in both the putative mirror neuron system and the sJA, but results were virtually identical to those of the peaks, showing that the peaks were indeed representative of the activity in clusters, and we therefore only report the results of the peak analyses.

Experiment II

The general procedures were very similar in Experiments I and II. In the interest of space, we will restrict ourselves to the differences between the methods below.

Participants

8 healthy volunteers (all right-handed; 5 female and 3 male; mean age 23.5 years ranging 21–24 years) from Experiment I.

Stimuli

Movies of a virtual game box replace the game box of Experiment I. The pictures of the custom-made MRI-compatible response box that we used for the first experiment were presented to the participants via a data projector on a screen that the participant could view through a mirror. Stimuli were programmed and presented using Presentation software (Neurobehavioral systems, Davis, CA).

At the beginning of each trial, an index finger holding the edge of the lower stick appeared. After that, the participants controlled that virtual finger using an MR compatible joystick (fORP, Current Designs, Inc., Philadelphia, USA) with their right index finger. For the 'human agent conditions' an index finger holding the edge of the upper stick appeared and the experimenter used her joystick in the control room to control the finger (Figs. S1E, G). In the computer condition, no such finger appeared on the upper stick (Figs. S1F, H). Although the participant was led to believe that the computer controlled the upper stick in the computer condition, the experimenter actually controlled the stick, again using her joystick from the control room. The experimenter in the control room viewed a clone of the movies viewed by the participant. The critical manipulation was that in the human condition, the experimenter viewed both the upper and lower halves of that screen, and could therefore adjust her actions to those of the participant, as in Experiment I, while in the computer condition, the lower half was occluded, preventing her from reaching to the participants' actions. In the computer condition, the participant therefore had to coordinate his/her actions to those of the 'computer' to reach the target within 200 ms, but the experimenter did not adjust hers (one-way coordination), whilst in the human condition, the two agents mutually adapted their actions to each other, creating the social loop so characteristic of joint actions (mutual coordination). What differs is whether the experimenter coordinated her actions with those of the participant (mutual vs. one-way coordination). After the end of each trial, both players had to place their joystick back onto the middle position. At rest, participants saw the entire response box with the experimenter's stick pointed up (12:00 of an analog clock) and the participant's down (06:00).

Procedure

All participants were invited to participate in Experiment II, but only 8 accepted. Only the joint action conditions (angle and straight) were acquired, but while the participant played with a human agent and what he/she thought to be a computer. The timing of conditions was as in Experiment I. The actual task was as in Experiment I: to cooperate with a human agent or react to the computer to shape the two sticks of the response box in an angle or a straight line (see Figs. S1E–H). The experimenter and participant then had to reach the target location virtually simultaneously (within 200 ms of each other) to jointly win the trial. The experimenter varied her initial movement velocity from trial-to-trial in both the human and computer conditions. In the human condition, both agents then adjust their movements to the velocity of the other to meet their common goal, conveying a mutual feeling of cooperation. In the computer condition, the burden of the adjustment rested entirely with the participant, as the experimenter was blind to the movements of the participant (one-way coordination). At the end of the experiment, we debriefed the participants about the experiment.

We asked the following questions: Did you perceive the trials as games? Did you try to cooperate during the experiment? Was there a difference between human and computer? Which game did you find harder; angle or straight? Were you able to control the joystick? Participants indeed perceived the computer condition as more difficult, but performance did not differ significantly (human 69% correct; computer 73.5% correct, t-test, $p > 0.14$) demonstrating that the participants successfully dealt with the challenge. The experimenter took care during the computer condition to generate movements that were similar in complexity and total duration to those in the human condition, including decelerations and accelerations to simulate those that occurred in response to the participant's behavior in the human condition. The lack of significant differences in total movement duration for the experimenter in the two conditions (see Results) confirms the similarity of overall movement characteristics. What changes however, was that these movements were no longer contingent with those of the participant, and participants reported perceiving the difference between human and computer agents in Experiment II (e.g. "The computer never waited for me!", one of the participants declared).

Participants were familiarized with the experimental conditions and the joystick during a short training session performed outside of the scanner prior to the scanning. In this session, participants were introduced to the computer condition by showing them a prerecorded motion (computer moving the stick), which they would observe and engage in joint actions with it. This was in contrast to the human condition in which they could see the experimenter side-by-side with them, playing a number of joint action trials to reinforce the feeling of cooperation. Moreover, joystick calibration and training with an online feedback was performed in the scanner before the start of the experiment. All participants reported that they perceived the computer conditions as controlled by a computer whilst they truly felt that they were playing with the experimenter in the human agent conditions.

Data acquisition and preprocessing: as in Experiment I

General data analyses

Functional data were analyzed using a general linear model (GLM) separately for each participant but only for the peak voxel of the ROIs determined in Experiment I and specified in Fig. 1. This was done because 8 participants provide sufficient statistical power while controlling for family wise error in a small number of ROIs but not for a whole brain analysis. Examining the responses from all the trials in each block using the surface analysis of Experiment I revealed a significant main effect of agent in a 2 Agent × 12 ROIs repeated measurement ANOVA ($p < 0.04$, human > computer), but examining the time course of the responses aligned to the beginning of each block revealed that in all ROIs, the responses decreased over the 8 trials of the blocks. We therefore remodeled the data using two sets of predictors for each block: one for the first and one for the remaining 7 trials. This analysis is the one we present in the manuscript. Signals were then analyzed using the same procedure as in Experiment I (see section "Analysis of peak voxels"), but using a repeated measurement ANOVA with 12 ROIs, 2 Agents (human vs. computer) and 2 conditions (angle vs. straight). The absence of a main effect or interaction of condition (all $p > 0.18$) motivated us to sum activity in the two conditions (angle and straight) and use a 2 Agent × 12 ROI ANOVA instead to test the one-tailed prediction that areas recruited during joint actions should respond more to the human agent than the computer. Using the first event only slightly improved the significance of the main effect of agent (from $p < 0.04$ with all trials to $p < 0.013$ with the first trial only, ANOVA 12 ROIs × 2 Agents). Least significant difference post-hoc t-tests were used to test differences in individual ROIs using a cut-off of $p < 0.05$.

Results and discussion

Results

Experiment I

To examine the role played by the putative mirror neuron system in joint actions, we first localized the putative mirror neuron system (Gazzola et al., 2007b; Keysers and Gazzola, 2006) by inclusively masking the contrast *obs*-rest with *exe*-rest (see Fig. 1b in blue and Table S1). This revealed areas corresponding to those reported in the literature including premotor (BA6, BA44), parietal (SI, SII, PF, SPL) and high-level visual areas (Fijii et al., 2007; Fogassi et al., 2005; Gallese et al., 1996, 2004; Gazzola et al., 2006, 2007a,b; Grafton et al., 1996; Hamilton et al., 2007; Heyes, 2001; Iacoboni and Dapretto, 2006; Iacoboni et al., 1999; Keysers and Gazzola, 2006; Keysers et al., 2003; Keysers and Perrett, 2004; Kohler et al., 2002; Newman-Norlund et al., 2007b; Rizzolatti and Craighero, 2004; Rizzolatti et al., 1996; Umilta et al., 2001 for a list of abbreviations). We then mapped brain areas involved in the dynamic integration of action observation, execution and task requirements by searching for voxels in which brain activity during joint actions (*str* or *ang*) exceeds the sum of that during *exe*, *obs* and *snd*. Voxels satisfying this criterion will be referred to as superadditive voxels in joint actions or sJA voxels and Fig. 1B (green) and Table S2 show their location. A central finding of this analysis is that sJA voxels do not consistently fall within the putative mirror neuron system but are adjacent to it. In the frontal lobe, sJA clusters were anterior to those of the putative mirror neuron system while in the parietal lobe, the sJA clusters were posterior to those of the putative mirror neuron system. Indeed, voxels common to both networks were rare and restricted to the superior parietal lobe and higher-level visual areas (see Fig. 1B in red and Table S3). Examination of the parameter estimates (Fig. 1C) of putative mirror neuron system and sJA peak locations shows that although the pattern of activity is somewhat similar, a functional dissociation exists: in putative mirror neuron system peak locations *str* and *ang* activity does not exceed *obs* + *exe* + *snd*, and the putative mirror neuron system therefore does not show evidence of *additional* processes during the integration of *obs*, *exe* and *snd* during joint actions. In sJA peak locations on the other hand, the activity is never significant during both *exe* and *obs*, showing that these areas are not part of the putative mirror neuron system. This does not mean however that the putative mirror neuron system is not involved in joint actions given that the putative mirror neuron system regions were significantly activated during *ang*, *str* as well as *obs* and *exe*, but that the putative mirror neuron system was not involved in the additional integration process.

Given that in most sJA regions, the *snd* condition determines a reduction of BOLD compared to baseline, we examined the voxels satisfying a definition of sJA excluding *snd* (i.e. *joint action* > *exe* + *obs*; see Fig. S3, Table S6 and triangles in Fig. 1C). This revised definition (sJA′) leads to very similar findings in frontal and parietal but not in the occipito-temporal region (around p11 in Fig. 1B) where listening to auditory instructions with closed eyes (*snd*) may have drawn attentional resources away from visual areas, and artificially inflated the contrast of joint actions against the control conditions. The overlap between sJA′ and putative mirror neuron system is however not larger than that between sJA and putative mirror neuron system. These control analyses strengthen the findings in frontal and parietal sJA, but commands care in interpreting the function of occipito-temporal sJA.

Finally, it has been argued that brain regions can be involved in integrating two modalities without the response to the multimodal stimulus exceeding the sum of its unimodal components (Beauchamp, 2005). An alternative, and sometimes more sensitive criterion, may be to request that the response to the multimodal stimulus exceeds the highest of its unimodal components (Beauchamp, 2005). Applying this maximum requirement to our data (see Materials and methods and Fig. S5) however leads to results that differed very little from

those of the previous analyses: none of the frontal, and only small regions of the parietal and temporal putative mirror neuron system clusters showed overlap with the voxels showing evidence for integration during joint actions. Indeed, in our particular data set, this maximum criterion was more conservative than the super-additivity criterion. Examining the parameter estimates of Fig. 1 helps understand why: in many sJA regions, one of the solo conditions was associated with negative parameter estimates, making the sum of the solo conditions inferior to their maximum.

In sum, all our analyses provide evidence that a network of brain regions including the left dorsal precentral gyrus shows evidence of integration during joint actions. However all these analyses also show that there is no overlap between regions showing evidence of integration and the putative mirror neuron system in the frontal lobe, and only restricted overlap in the parietal lobe and higher-level visual areas. Although one might argue that relaxing statistical thresholds or increasing the statistical power of the experiment might reveal overlaps between these networks in the frontal lobe, our results do suggest that the voxels most reliably associated with integration and the putative mirror neuron system differ.

Furthermore, as mentioned in the Introduction, unlike imitation in the strict sense (Thorpe, 1956), in which the rule that links observed and executed actions is constant ('You do X so I do X'), joint actions often require changing this rule (Heyes, 2001; Iacoboni and Dapretto, 2006; Iacoboni et al., 1999; Newman-Norlund et al., 2007b). Relocating a dinner table for instance can involve changing from moving it sideways ('You move North so I move North') to turning it around ('You move North so I move South'). The integrative component of joint actions could therefore be split in two subprocesses: determining which rule is appropriate at a certain moment in time and then implementing this rule within the perception–action loop (Newman-Norlund et al., 2007a; Sebanz et al., 2006a). Given that our definition of joint actions is based on the *str* and *ang* blocks in which the rule stays constant across the 8 trials of a block, this definition will mainly capture voxels involved in implementing the rule, because processes involved in determining the rule would only occur once during a block and have a weak impact on the overall block activity. To capture brain areas involved in determining the rule, we additionally compared brain activity in the *mix* condition with the unmixed joint action blocks (contrast: $2mix > str + ang$). Fig. S4 and Table S4 show that regions augmenting their blood flow when rules have to be changed more frequently (yellow) overlap with both putative mirror neuron system and sJA in the parietal (pink and brown, including part of the small overlap between putative mirror neuron system and sJA) but not the frontal lobe, pointing towards a functional dissociation between parietal and frontal nodes of both the putative mirror neuron system and sJA: while the frontal regions appear involved primarily in implementing the rule, the parietal regions seem also to participate in determining this rule.

Experiment II

While Experiment I determined a number of brain areas involved in integrating the observation of an external event with the execution of one's own actions, it cannot determine if this integration reflects joint actions in the strict sense, i.e. the mutual coordination of two agents. To examine this question, we scanned 8 participants again using a modified version of the game. Participants now used a joystick to manipulate a representation of the game box on the screen. This allowed us to contrast two conditions: (a) the participant played with a human agent that reacted to the participants own actions as in Experiment I and (b) the participant played with what he believed to be a computer, and which did not react to the participant's actions (see Materials and methods and Figs. S1E–H).

Debriefing of the participants after Experiment II confirmed that they felt they indeed played with the experimenter in condition (a)

and with a computer in condition (b). They additionally commented that the experimenter was "friendlier" than the computer: only the human agent was perceived as cooperative whilst the computer, which they said "never waited for them", was not. All felt that using a joystick to control the game box made the task more difficult than using the 'real' game box of Experiment I. The average duration of the movement however was not different in the two conditions (2263 ms in the human and 2115 ms in the computer condition, *t*-test, $p > 0.42$). The proportion of correct trials (i.e. the experimenter and participant reaching the correct target location within 200 ms of each other) was lower in the first session (human condition: 53% correct, computer condition: 56% correct), reflecting the initial difficulty in controlling the game box with the joystick, but improved in the remaining sessions, arriving at 69% correct overall for the human condition and 73.5% for the computer condition. Importantly, there was no significant difference between the performance in the two conditions (*t*-test, two-tailed, $p > 0.14$).

We extracted in the peak voxel of the 12 ROIs identified in Experiment I (Fig. 1), the activity during the straight and angle conditions while participants played with the human agent and the computer, and analyzed the results using an ANOVA with 2 conditions (*str* vs. *ang*), 12 ROIs and 2 agents (human vs. computer). The main effect of Condition was not significant ($p > 0.15$), nor did Condition interact with the other factors (all $p > 0.09$). Reanalyzing the data using the sum of angle and straight in a 12 ROI × 2 Agents ANOVA revealed a main effect of Agent ($p < 0.013$) with activity while participant playing with the human agent being higher than when playing with a computer. In addition, there was a main effect or ROI ($p < 10^{-7}$), and an interaction of ROI × Agent ($p < 0.001$). Post-hoc comparison revealed that although activity was numerically larger for the human condition in all the ROIs of both the putative mirror neuron system and sJA, this difference was significant for dorsal frontal sJA (ROI3), ventral putative mirror neuron system (ROI2), the high-level visual (ROIs 10 and 11) and many of the parietal regions (ROIs 6, 7, and 8; see Fig. 2, all $p < 0.05$, one-tailed LSD-post-hoc test).

In summary, both the putative mirror neuron system and sJA network were more responsive while playing with a human agent that responded to the actions of the participant, compared to playing with a computer that did not. This difference cannot be explained by the participants paying less attention to the actions of the computer, as the number of correct trials did not differ in the two conditions, nor can it be explained by differences in the time spent moving with the human agent or the computer, as the playing time did not differ significantly ($p > 0.42$).

Discussion

Our aims were to identify the circuitry specifically involved during the task-dependent *integration* of observed and executed actions that distinguishes joint actions from the action observation and execution done in isolation. In particular, we also aimed to examine the degree to which this process occurs within or beyond the putative mirror neuron system. In Experiment I, we found evidence for a distributed network of brain areas showing additional activity during joint actions compared to both the sum of solo observation and execution of similar actions (with or without taking the *snd* condition into account) and the maximum of solo observation and execution. As we predicted, this joint action network however overlaps remarkably little with the putative mirror neuron system, and not at all within the frontal lobe. In Experiment II, we found that in the joint actions and putative mirror neuron system networks, activity was stronger while participants played with a human agent that reacted to their own actions than while playing with a computer that did not.

In the following, we will first discuss the potential functional contribution of the main joint action clusters during Experiment I, and will assess critically how these findings constrain the role played by
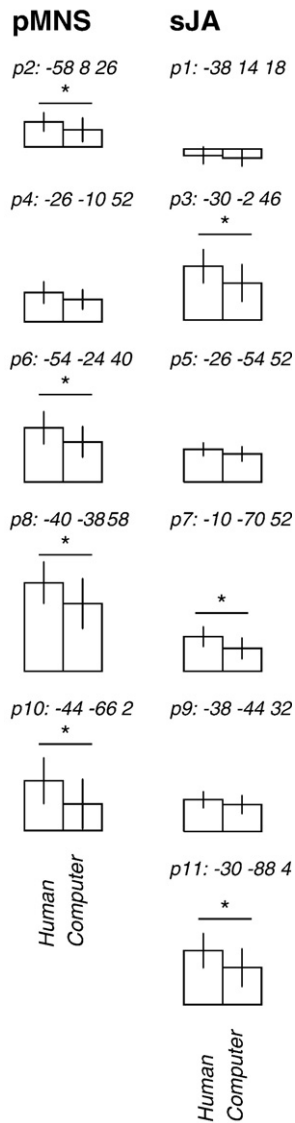
Fig. 2. Parameter estimates of the peak locations of Experiment I during joint action with a human (human) and with a computer (computer) in Experiment II. Peaks are numbered as in Fig. 1C (numbers indicating the MNI coordinates of the peak). All parameter estimates and the error bars (SEM) are shown on the same scale and can be directly compared (actual units irrelevant because arbitrary). *: Significant one-tailed human > computer differences according to LSD post-hoc comparison.

the other individual's actions may have heightened selective visual attention to both agents' actions during joint actions, causing part or all of the enhanced BOLD response in these regions. Such visual attention however would not be an epiphenomenon, but a functionally important mechanism to ensure optimal visual processing for action. Fig. S3 shows that excluding the *snd* condition from the definition of sJA limits but does not abolish the involvement of these regions in joint actions, suggesting that the contribution of high-level visual areas to joint actions, be it integrative and/or attentional in nature, is genuine.

Second, the putative mirror neuron system is known to transform the vision of actions into motor representations of similar actions (Etzel et al., 2008; Gallese et al., 2004; Gazzola and Keysers, 2009; Iacoboni and Dapretto, 2006; Keysers and Gazzola, 2006; Liepelt et al., 2008b; Rizzolatti and Craighero, 2004). Accordingly, it represents both partners' actions in a common code (Prinz, 1997) that is probably motor in the premotor regions but could be somatosensory or visual in other regions of the putative mirror neuron system (Gazzola and Keysers, 2009). In our task, during joint actions, the activity of the putative mirror neuron system was simply the sum of its activity during *obs + exe* or *obs + exe + snd*. This finding suggests that the activity observed in the putative mirror neuron system during joint actions appears to reflect two additive processes. During execution, activity in premotor and parietal regions probably reflects motor planning, while in high-level visual areas it may reflect the transformation of motor plans in the expected visual consequences of these actions using forward models. During observation, activity in high-level visual areas probably reflects processing of the visual stimulus itself while parietal and premotor activations would reflect activation of corresponding motor plans through inverse models (Gazzola and Keysers, 2009). Experiment I shows that in these regions, there is however no evidence for any additional, integrative processing between the streams of information corresponding to the two agents' actions. Neighboring areas of the sJA network may instead be responsible for these additional processes. This finding apparently contrasts with reports showing that the "mirror neuron system is activated to a greater extent during execution of actions carried out with a partner as compared to actions carried out alone" (Newman-Norlund et al., 2008). These authors however did not map the putative mirror neuron system in their participants, and their IFG clusters (47, 16, and 25) fell outside of the putative mirror neuron system in our study.

A challenge for motor control during joint actions is the fact that our visual and motor systems have relatively long latencies (several hundreds of milliseconds). Our actions would thus lag behind those of our partner. An interesting property of the mirror neuron system however is that it is known to anticipate future actions that are not yet fully visible (Umilta et al., 2001). Motor control would then not react to an outdated representation of what the other person did several hundreds of milliseconds ago, but to an anticipation of his future actions, and the mirror neuron system would thereby contribute to solve this time lag issue (Kilner et al., 2004; Urgesi et al., 2006). The lack of additional activity in the putative mirror neuron system during joint actions compared to solo actions suggests that this anticipation is triggered spontaneously both during solo observation and during joint actions.

Third, a large network including the posterior parietal lobe (inferior and superior parietal lobule; precuneus), the basal ganglia and cerebellum showed enhanced activity in joint actions. These areas are known to act in concert when monkeys and humans learn and use novel sensory–motor associations whether they involve other biological agents or not (Gold and Shadlen, 2007; Hikosaka et al., 2002). This network could therefore be crucial for transforming the novel and changing task requirements of our game into an appropriate way to map observed onto executed actions. The increase of activity in the mixed condition in the parietal location of this network would

the putative mirror neuron system in our joint action task. Thereafter, we will discuss how Experiment II suggests that this activity is influenced by the presence of a social loop.

First, high-level visual areas, including locations in the vicinity of the EBA (Downing et al., 2001) and STS (Puce and Perrett, 2003), are known to respond preferentially to the vision of biological agents and actions but also during blind action execution (Astafiev et al., 2004; Gazzola et al., 2006, 2007a,b; Gazzola and Keysers, 2009; Iacoboni et al., 2001). The presence of these regions in our sJA and maxJA networks suggests that the process of integrating observed and executed actions may not only occur in frontal regions employing a motor code. Instead, this integration may also occur at a more sensory level. As suggested by the idea of forward models in motor control (Gazzola and Keysers, 2009; Wolpert et al., 2003; Wolpert and Ghahramani, 2000), the intended actions of the participant could be transformed into expected sensory consequences in high-level visual cortex which can then be compared and integrated with the observed actions of the other. Alternatively, the need to act based on observing

support the idea that this node is particularly involved in transforming the task requirement into an appropriate visuo-motor link.

Finally, the set of ventral and dorsal frontal clusters anterior to the putative mirror neuron system do not respond during execution or observation and therefore seem neither mirror nor premotor. The absence of responses during the vision of actions, and the absence of enhancement in the mixed compared to the unmixed conditions in these areas, make it unlikely that the preferential response during joint actions reflects the effect of visual attention or attentional memory load alone. Given that the rostral BA44 also contains regions responding during imitation but not observation or execution of finger movements (Molnar-Szakacs et al., 2005), this suggests that integrating two agents' actions recruits brain regions just anterior to the precentral and inferior frontal regions involved in motor control. This proximity would enable these regions to integrate very closely with the motor control and sensory functions ascribed to premotor areas in general and the putative mirror neuron system in particular. The detailed selectivity pattern of these regions and the nature of the computations they perform will however need to be addressed in future experiments.

By examining the results of Experiment I alone, we know that sJA areas are involved in the integration of a visual input and a motor output, but we cannot know if this integration activity is specific for the mutual coordination that defines joint actions or would be just as strong during a task only requiring one-way coordination. In Experiment II, we therefore acquired brain activity in the ROIs of Experiment I in half of our participants while they played the same cooperation game with either a human agent that reacted to their own actions (mutual coordination) or (with what they thought to be) a computer, that did not (one-way coordination). Importantly, throughout the sJA and putative mirror neuron system ROIs, activity was higher while playing with the human agent, as demonstrated by the significant main effect of agent. This shows that despite the presence of biological movement in both conditions (given that a human experimenter blind to the participant's actions was actually playing the role of the computer), the presence of a human finger in the display, the belief to be playing with a human agent and/or the contingency that participants detected between the human agent and their own actions (mutual coordination) must have made these networks sensitive to the presence of the social loop that characterizes joint actions (Liepelt et al., 2008a).

The task-dependent *integration* of action observation and execution during joint actions however would occur outside of the putative mirror neuron system, in accord with our hypothesis and the previous theoretical proposals (Newman-Norlund et al., 2007a).

In summary, a number of studies have suggested that the IFG is not only involved when we respond to the actions of others by doing the same as they do (imitation) but also when responding with complementary actions (Newman-Norlund et al., 2007b) or engaging in joint actions (Newman-Norlund et al., 2007a). This has lead to the idea that the putative mirror neuron system would be responsible for integrating one's own actions to those of others in joint actions. In contrast, we hypothesized that the flexibility required during joint actions goes beyond the known properties of the mirror neuron system. Supporting our hypothesis, we find brain regions to be involved in integrating observed and executed actions during the social loop of joint actions, but these regions are distinct from the putative mirror neuron system that is engaged during solo observation and execution. In contrast to recent claims, our data therefore suggest that joint action may be a dual process:

One set of areas (including the putative mirror neuron system) seems to 'simply' transform observed actions into representations of similar actions in the observer through a combination of forward and inverse models (Gazzola and Keysers, 2009). This ensures that the two essential components that need integration during joint actions are in the same neural code: our own actions and those of others. This code

can be relatively motor, sensory or hybrid in different regions of the brain (premotor, STS or parietal), and the translation between these codes could depend on the forward and inverse models we build up while observing the consequences of our own actions and preparing the participant to act. This is compatible with the existing the animal literature (Fogassi et al., 2005; Gallese et al., 1996; Keysers et al., 2003; Kohler et al., 2002). In the monkey, this transformation predicts the goal of observed actions (Umilta et al., 2001), providing the observer with the opportunity to tune his actions to the expected actions of others instead of 'lagging behind' due to the latencies of the visual and motor systems. Premotor neurons similar to the minority of 'logically connected' neurons in the macaque could serve to ensure that the participant now has a number of actions primed in his brain: actions both similar and complementary to those of the other individual.

The second set of areas showing additional activity during integration in joint actions then utilizes these common codes and behavioral alternative to integrate flexibly our own actions with those of others and select the most adequate action, amongst the alternatives primed in the premotor cortex, to achieve our current joint goals.

Further research investigating the functional connections between the two networks will shed more light on the roles of these distinct, but probably communicating, networks in one of our most defining features: our capacity to cooperate constructively with other members of our species.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2009.06.010.

References

Astafiev, S.V., Stanley, C.M., Shulman, G.L., Corbetta, M., 2004. Extrastriate body area in human occipital cortex responds to the performance of motor actions. Nat. Neurosci. 7, 542–548.
Beauchamp, M., 2005. Statistical criteria in fMRI studies of multisensory integration. Neuroinformatics 3, 93–114.
Blakemore, S.J., Bristow, D., Bird, G., Frith, C., Ward, J., 2005. Somatosensory activations during the observation of touch and a case of vision–touch synaesthesia. Brain 128, 1571–1583.
Catmur, C., Walsh, V., Heyes, C., 2007. Sensorimotor learning configures the human mirror system. Curr. Biol. 17, 1527–1531.
Chong, T.T., Cunnington, R., Williams, M.A., Kanwisher, N., Mattingley, J.B., 2008. fMRI adaptation reveals mirror neurons in human inferior parietal cortex. Curr. Biol. 18, 1576–1580.
Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.
Etzel, J.A., Gazzola, V., Keysers, C., 2008. Testing simulation theory with cross-modal multivariate classification of fMRI data. PLoS ONE 3, e3690.
Fijii, N., Hihara, S., Iriki, A., 2007. Social cognition in premotor and parietal cortex. Social Neurosci. 3.
Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. Science 308, 662–667.
Friston, K.J., Penny, W.D., Glaser, D.E., 2005. Conjunction revisited. NeuroImage 25, 661–667.
Gallese, V., 2003. The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. Psychopathology 36, 171–180.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119 (Pt 2), 593–609.

Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. 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, 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), 1239–1255.

Gazzola, V., Aziz-Zadeh, L., Keysers, C., 2006. Empathy and the somatotopic auditory mirror system in humans. Curr. Biol. 16, 1824–1829.

Gazzola, V., Rizzolatti, G., Wicker, B., Keysers, C., 2007a. The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. NeuroImage 35, 1674–1684.

Gazzola, V., van der Worp, H., Mulder, T., Wicker, B., Rizzolatti, G., Keysers, C., 2007b. Aplasics born without hands mirror the goal of hand actions with their feet. Curr. Biol. 17, 1235–1240.

Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.

Grafton, S.T., Arbib, M.A., Fadiga, L., Rizzolatti, G., 1996. Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. Exp. Brain Res. 112, 103–111.

Hamilton, A.F., Brindley, R.M., Frith, U., 2007. Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? Neuropsychologia 45, 1859–1868.

Heyes, C., 2001. Causes and consequences of imitation. Trends Cogn. Sci. 5, 253–261.

Hikosaka, O., Nakamura, K., Sakai, K., Nakahara, H., 2002. Central mechanisms of motor skill learning. Curr. Opin. Neurobiol. 12, 217–222.

Iacoboni, M., Dapretto, M., 2006. The mirror neuron system and the consequences of its dysfunction. Nat. Rev., Neurosci. 7, 942–951.

Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G., 1999. Cortical mechanisms of human imitation. Science 286, 2526–2528.

Iacoboni, M., Koski, L.M., Brass, M., Bekkering, H., Woods, R.P., Dubeau, M.C., Mazziotta, J.C., Rizzolatti, G., 2001. Reafferent copies of imitated actions in the right superior temporal cortex. Proc. Natl. Acad. Sci. U. S. A. 98, 13995–13999.

Jabbi, M., Swart, M., Keysers, C., 2007. Empathy for positive and negative emotions in the gustatory cortex. NeuroImage 34, 1744–1753.

Keysers, C., Gazzola, V., 2006. Towards a unifying neural theory of social cognition. Prog. Brain Res. 156, 379–401.

Keysers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. Trends Cogn. Sci. 8, 501–507.

Keysers, C., Kohler, E., Umilta, M.A., Nanetti, L., Fogassi, L., Gallese, V., 2003. Audiovisual mirror neurons and action recognition. Exp. Brain Res. 153, 628–636.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.L., Fogassi, L., Gallese, V., 2004. A touching sight: SII/PV activation during the observation and experience of touch. Neuron 42, 335–346.

Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.J., Sirigu, A., 2004. Motor activation prior to observation of a predicted movement. Nat. Neurosci. 7, 1299–1301.

Knoblich, G., Jordan, S., 2002. The mirror system and joint action. In: Gallese, M.I.S.V. (Ed.), Mirror Neurons and the Evolution of Brain and Language. John Benjamins, Amsterdam, pp. 115–124.

Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., Rizzolatti, G., 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.

Liepelt, R., Cramon, D.Y., Brass, M., 2008a. What is matched in direct matching? Intention attribution modulates motor priming. J. Exp. Psychol. Hum. Percept. Perform. 34, 578–591.

Liepelt, R., Von Cramon, D.Y., Brass, M., 2008b. 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 43, 784–792.

Molnar-Szakacs, I., Iacoboni, M., Koski, L., Mazziotta, J.C., 2005. Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. Cereb. Cortex 15, 986–994.

Newman-Norlund, R.D., Noordzij, M.L., Meulenbroek, R.G.J., Bekkering, H., 2007a. Exploring the brain basis of joint action: coordination of actions, goals and intentions. Social Neurosci. 2, 48–65.

Newman-Norlund, R.D., van Schie, H.T., van Zuijlen, A.M., Bekkering, H., 2007b. The mirror neuron system is more active during complementary compared with imitative action. Nat. Neurosci. 10, 817–818.

Newman-Norlund, R.D., Bosga, J., Meulenbroek, R.G., Bekkering, H., 2008. Anatomical substrates of cooperative joint-action in a continuous motor task: virtual lifting and balancing. NeuroImage 41, 169–177.

Prinz, W., 1997. Perception and action planning. Eur. J. Cogn. Psychol. 9, 129–154.

Puce, A., Perrett, D., 2003. Electrophysiology and brain imaging of biological motion. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 435–445.

Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends Neurosci. 21, 188–194.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., Fazio, F., 1996. Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp. Brain Res. 111, 246–252.

Sebanz, N., Bekkering, H., Knoblich, G., 2006a. Joint action: bodies and minds moving together. Trends Cogn. Sci. 10, 70–76.

Sebanz, N., Knoblich, G., Prinz, W., Wascher, E., 2006b. Twin peaks: an ERP study of action planning and control in co-acting individuals. J. Cogn. Neurosci. 18, 859–870.

Sebanz, N., Rebbechi, D., Knoblich, G., Prinz, W., Frith, C.D., 2007. Is it really my turn? An event related fMRI study of task sharing. Social Neurosci. 2, 81–95.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D., 2004. Empathy for pain involves the affective but not sensory components of pain. Science 303, 1157–1162.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D., 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439, 466–469.

Thioux, M., Gazzola, V., Keysers, C., 2008. Action understanding: how, what and why. Curr. Biol. 18, R431–R434.

Thorpe, W., 1956. Learning and Instinct in Animals. Methuen, London.

Umilta, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., Rizzolatti, G., 2001. I know what you are doing. a neurophysiological study. Neuron 31, 155–165.

Urgesi, C., Moro, V., Candidi, M., Aglioti, S.M., 2006. Mapping implied body actions in the human motor system. J. Neurosci. 26, 7942–7949.

Wicker, B., Keysers, C., Plailly, J., Royet, J.P., Gallese, V., Rizzolatti, G., 2003. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron 40, 655–664.

Wolpert, D.M., Ghahramani, Z., 2000. Computational principles of movement neuroscience. Nat. Neurosci. 3 Suppl, 1212–1217.

Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 593–602.