

Neural mechanisms and models underlying joint action

Fabian Chersi

Received: 14 December 2010 / Accepted: 5 April 2011 / Published online: 28 April 2011
© Springer-Verlag 2011

Abstract Humans, in particular, and to a lesser extent also other species of animals, possess the impressive capability of smoothly coordinating their actions with those of others. The great amount of work done in recent years in neuroscience has provided new insights into the processes involved in joint action, intention understanding, and task sharing. In particular, the discovery of mirror neurons, which fire both when animals execute actions and when they observe the same actions done by other individuals, has shed light on the intimate relationship between perception and action elucidating the direct contribution of motor knowledge to action understanding. Up to date, however, a detailed description of the neural processes involved in these phenomena is still mostly lacking. Building upon data from single neuron recordings in monkeys observing the actions of a demonstrator and then executing the same or a complementary action, this paper describes the functioning of a biologically constraint neural network model of the motor and mirror systems during joint action. In this model, motor sequences are encoded as independent neuronal chains that represent concatenations of elementary motor acts leading to a specific goal. Action execution and recognition are achieved through the propagation of activity within specific chains. Due to the dual property of mirror neurons, the same architecture is capable of smoothly integrating and switching between observed and self-generated action sequences, thus allowing to evaluate multiple hypotheses simultaneously, understand actions done by others, and to respond in an appropriate way.

Keywords Joint action · Mirror system · Motor chains · Intention inference · Computational model

Introduction

Human beings are capable of very sophisticated forms of inferential processes, planning and action coordination. In collaborative tasks, individuals effortlessly monitor the actions of their partners, interpret them in terms of their outcomes, and use these predictions to select an adequate complementary behavior (Sebanz et al. 2006; Sebanz and Knoblich 2009).

A consistent amount of studies have demonstrated that animals are also able to engage in various types of joint behavior that involve some form of cooperation and coordination among individuals (Noe 2006; Grinnel et al. 1995; Chalmeau 1994; Melis et al. 2006) [although, the question about how deeply animals understand the role and intentions of collaborative partners is still topic of contrasting debate (Visalberghi et al. 2000)].

Taken together, these results lead to think that there exist a common low-level neural mechanism or structure, more complex and developed in humans and less in animals, which underlies social interaction and joint action.

In a series of key experiments, researchers at the laboratory of Rizzolatti (Di Pellegrino et al. 1992; Gallese et al. 1996; Rizzolatti et al. 2001) discovered that a consistent percentage of neurons in the premotor cortex (area F5) become active not only when monkeys execute purposeful object-oriented motor acts, such as grasping, tearing, holding, or manipulating objects, but also when they observe the same actions executed by another monkey or even by a human demonstrator. This type of neurons have later been termed “mirror neurons” to underlie their

F. Chersi (✉)
Istituto di Scienze e Tecnologie della Cognizione,
CNR, Via S. Martino della Battaglia 44, Roma, Italy
e-mail: fabian.chersi@istc.cnr.it

capacity to respond to the actions of the others as if they were the reflection of one owns.

Neurons with the same mirroring properties, i.e., matching the observation and the execution of actions, have been subsequently discovered also in the inferior parietal lobule (IPL) of the monkey, more precisely in areas PF and PFG (Gallese et al. 2002; Fogassi et al. 2005). Traditionally considered only an association cortex, it is now clear that the parietal cortex is actively involved in the execution and interpretation of motor actions.

An important distinction between the two mirror areas, which at first sight, may seem to be just one the replication of the other, is that while F5 is considered to contain the hand motor vocabulary (Gentilucci and Rizzolatti 1990), i.e., it codes detailed movements such as “precision grip”, “finger prehension”, or “whole hand grasp”, in IPL, the motor act representation is more abstract with neurons encoding a generic “grasp” or “reach” or “place”, rarely including parameters such as speed or force. In this view, it appears that the parietal cortex provides the premotor cortex, with which it is strongly interconnected, with high-level instructions, which are then transformed into more concrete (i.e., low-level) motor commands with F5 integrating, for example, details about object affordances (provided by neurons in the intraparietal sulcus) and the primary motor cortex resolving the correct muscle synergies. It is important to note that, in both these areas, there is a strong modulation of the neuronal responses by the overall goal of the action (Fogassi et al. 2005; Fogassi et al. 2007; Bonini et al. 2010).

One fundamental question concerns the source of the visual input to mirror neurons. There is a striking resemblance between the visual properties of mirror neurons and those of a class of neurons present in the monkey’s (Perrett et al. 1989; Oram and Perrett 1994) anterior part of the superior temporal sulcus (STS). Neurons in this area, which is strongly connected to the IPL, are also selectively responsive to hand–object interactions, such as reaching for, retrieving, manipulating, picking, tearing, and holding but they are not mirror neurons, as they do not discharge during action execution. The existence of these neurons is of particular importance because it indicates that mirror neurons do not directly perform visual processing but they receive an already view-invariant description of the interaction between effectors and objects observed in a scene. In other words, the signals arriving to PF/PFG correspond to the motor content extracted from the visual input. In this sense, the transmission of information from STS to IPL has the effect of “translating” it into a “motor format”, which is the same as that used by motor neurons. The activation of mirror neurons is thus the result of the combination of inputs deriving from temporal areas and from motor neurons in same or connected areas (corollary discharges).

In the last decades, a great number of brain imaging studies and electrophysiological experiments have provided evidence for the existence of a mirror neuron system (MNS) in the human brain (for a review see (Gallese et al. 2004; Rizzolatti and Craighero 2004)). In particular, the link between STS and IPL (see Fig. 1) has been demonstrated to exist also in humans (Frith and Frith 1999; Allison et al. 2000; Grossman et al. 2000), as well as the circuit between premotor areas and parietal areas, both directly (Rushworth et al. 2006) and indirectly (Iacoboni et al. 2001; Iacoboni et al. 2005).

Furthermore, it has been shown that, also in humans, the observation of actions made by others elicits a sequential pattern of activation in the MNS that is consistent with the proposed pattern of anatomical connectivity (Nishitani and Hari 2000; Buccino et al. 2001).

The key aspect of interest for this study is, at this point, how the mirror neuron system may be of benefit for individuals during joint actions. The position sustained in this study is that the MNS plays a fundamental role in low-level visual and motor understanding, thus intervening in processes that do not require agentive or meta-representational understanding. This may be the case, for example, when executing habitual or overtrained actions or for responses that require high speed and low reaction times. Finally, most probably, this is the only mechanism available to less evolved creatures that do not possess highly cognitive brain mechanisms that allow for complex simulation and mind reading capabilities.

The approach adopted in the rest of the article will be to initially provide neurophysiological data, showing the sequential and specific activation of mirror neurons during both action execution and action observation tasks, then

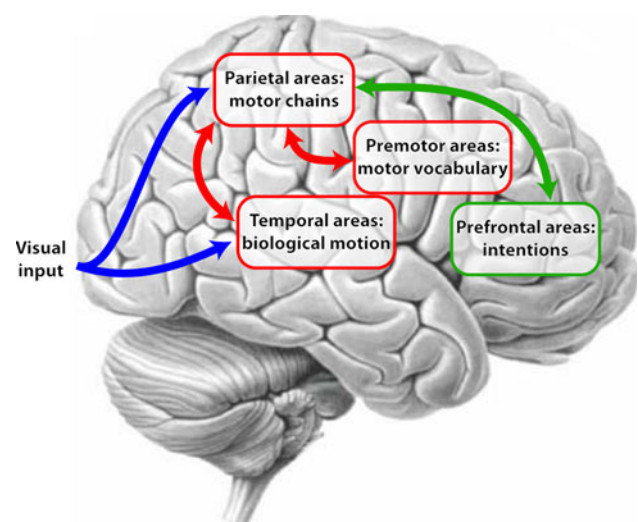


Fig. 1 Schematic representation of the mirror neuron system (red) with connections to the prefrontal cortex (green) and visual areas (blue)

to describe the reference experiment in which the above-mentioned tasks, alternatively executed by two agents, will be assembled in order to obtain a new joint action task.

The focus of the modeling part will be to describe a neural architecture in which action sequences are encoded as chains of mirror neuron subpopulations encoding single motor acts allows to easily track the actions of other individuals and to execute corresponding response actions. Additionally, it will show that exploiting the connections between the parietal mirror neurons and the prefrontal cortex, it is possible to obtain a mechanism that constantly evaluates the actions of others and ascribes them motor intentions. A key aspect of this model, compared with more classical approaches, is that thanks to the dual property of mirror neurons, there is no need for separate modules for action observation and execution, which, as will be explained later, presents some considerable advantages.

Materials and methods

Experimental setup

The recordings of mirror neurons' activity presented in this paper are the result of a new analysis, aimed at highlighting the temporal relations, of the data partially described in the study by (Fogassi et al. 2005; Chersi et al. 2005). All neurons were recorded in the rostral part of IPL of two behaving monkeys (*M. nemestrina*). The studied units were active in association with movements of the hand/arm, (mostly grasping but also reaching, placing, and bringing to the mouth) and were tested in two modalities. In the "motor condition", the monkey starting with the hand in a fixed "home" position reached for and grasped a piece of food located in front of it or occasionally from the experimenter's hand and brought it to the mouth, and alternatively, it reached for and grasped a metal cube in the same position and placed it into a container nearby. In the "visual condition", the monkey had to observe the experimenter executing the two action sequences described earlier. A detailed description of the experimental setup can be found in the study by Fogassi et al. (2005b).

Interestingly, the firing rate of the large majority (around 70%) of motor and mirror neurons was influenced by the final goal of the action. More specifically, neurons that were highly active during the grasping-to-eat action fired only weakly when the goal was placing the metal cube into the container and viceversa. Figure 2 shows the activity of three IPL mirror neurons during motor execution (left panel) and during observation (right panel) of the "reaching, grasping food, bringing to the mouth" action sequence. As can be seen, this graph clearly indicates that they encode

different phases of the action and that their representation is highly congruent across the two modalities.

To exclude the possibility that the differential discharge was due to characteristics of the two objects or to the hand trajectories, the monkeys were trained to place the piece of food (by giving a tastier reward) and the metal cube in a container near to the mouth. It was found that neuron selectivity did not depend on the physical characteristics of the presented stimulus nor on the position in space. All together, these results can be summarized by saying that the majority of motor and mirror neurons in the inferior parietal lobe encode single motor acts in an abstract way, that is, they do not represent physical characteristics of objects nor their location in space or the speed of the movement. In contrast, their response is strongly modulated by the final goal of the action (Fogassi et al. 2005).

The joint action tasks that are modeled in this study have been obtained by virtually combining the "visual" and "motor" conditions described earlier. More precisely, in the new setup, two monkeys sit at a table facing each other at a reaching distance.

In the first new condition, the first individual is given a piece of food but has to execute a "grasping to place" toward the table center. At the same time the second individual observes the action and then produces a "grasping to eat" sequence. In the second new condition, the first individual is given a metal cube and has to execute again a "grasping to place". The second individual observes the scene and also produces a "grasping to place".

As can be easily inferred, the new tasks produce a simultaneous and/or subsequent activation of mirror and motor circuits. (A more detailed description is provided below). The neural responses and the considerations reported earlier are valid also for the new setup.

The following sections of the paper will concentrate on modeling and reproducing the neuronal activity recorded in the brain of the second individual.

Modeling the mirror system

It is generally accepted that the mirror circuit plays a very important role in action recognition in both monkeys and humans (Rizzolatti et al. 2001). However, the detailed mechanisms of this process remain still relatively unclear. In recent years, there has been an increasing interest in modeling the mirror neuron system and action recognition mechanisms (Fagg and Arbib 1998; Oztop and Arbib 2002; Bonaiuto et al. 2007) and sensory-motor couplings (Demiris and Hayes 2002; Haruno et al. 2001; Craighero et al. 2007). In particular, the Chain Model (Chersi et al. 2005; Chersi et al. 2006) addressed these questions in detail at neuronal level. Its key hypothesis is that action sequences are encoded in IPL as chains of neurons that

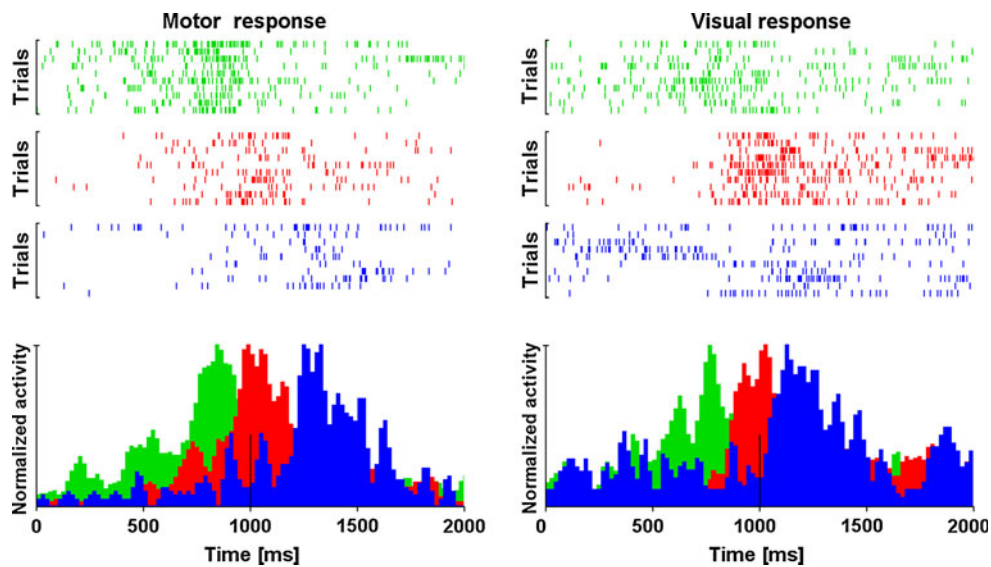


Fig. 2 Activity of three mirror neurons recorded in the monkey IPL during the execution of a “grasping to eat” sequence (*left panel*) and during the observation of an experimenter performing the same action (*right panel*). Rasters and histograms were synchronized with the moment in which the monkey or the experimenter touched the piece

of food ($t = 1,000$ ms). Different colors indicate different motor acts: *green* reaching, *red* grasping, and *blue* bringing to the mouth. The identification of the encoded act was achieved by comparing activity timings with sensors’ signals. Neuronal discharge frequency has been normalized for comparison

represent subsequent motor acts leading to the desired goal. According to this view, the action of taking a piece of food, for example, is encoded as the concatenation of neurons that represent the reaching, the grasping, and the retrieving motor act. The execution of an action corresponds to the propagation of an activity wave within the corresponding neural chain triggered by external input from PFC or from sensory areas. Importantly, due to their “duality”, the same mechanism for action execution is employed for recognition in chains composed of mirror neurons. More precisely, during the execution of actions, mirror neurons behave exactly as ordinary motor neurons, while during observation specific pools of mirror neurons resonate when the observer sees the corresponding motor acts (e.g., reaching, grasping, placing) being executed by another individual (see Fig. 3).

Model hypotheses

Joint actions require continuous tracking and interpretation of other’s movements, as well as the selection and execution of corresponding motor plans. Based on the neurophysiological data reported in the previous sections, the following mechanisms were hypothesized to underlie these capabilities in this model.

- The execution and the interpretation of actions utilize the same neural circuit, i.e., the mirror neuron system. The different outcome derives from the fact that during observation, neuronal activity does not propagate to motor areas due to a low-level blocking mechanism

involving areas such as the supplementary motor areas (SMA) and the basal ganglia (BG).

- The core of the implemented circuit comprises the parietal cortex that contains neuronal chains encoding goal-directed motor sequences, and the prefrontal cortex that encodes intentions and task-relevant information. Sensory and temporal areas provide the necessary input to trigger the propagation of activity within specific chains.
- When an observed action elicits the execution of the same action (e.g., a “grasping to place” action recalls a “grasping to place” response), the exact same mirror chains are recruited.
- During observation, feedback signals from resonating mirror neurons increase the activity of specific neurons in PFC encoding intentions attributed to the observed individual.
- As long as there is no interference or cross-talk between neuronal chains, multiple hypotheses about the observed action can be evaluated simultaneously.
- When engaged in joint action, the prefrontal cortex controls the concatenation (but not the execution) of different actions chain.

Model description

The model proposed in this paper is based on the Chain Model described earlier but introduces some important modifications. The main components (see Fig. 4) are a

Fig. 3 Scheme representing the neuronal chains and the flow of information between an acting and an observing individual. This information is used by the observer first to produce a cognitive representation of the other's intention and then a corresponding motor response activating either the same or a different action chain

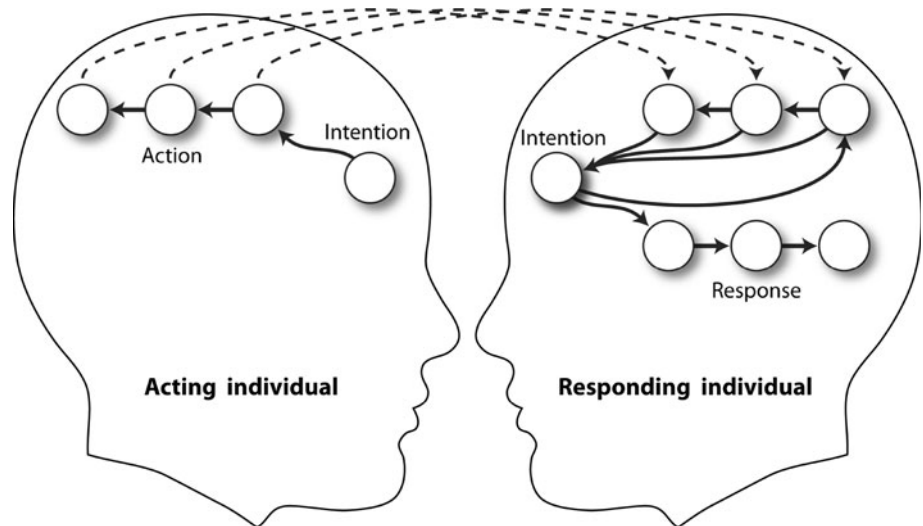
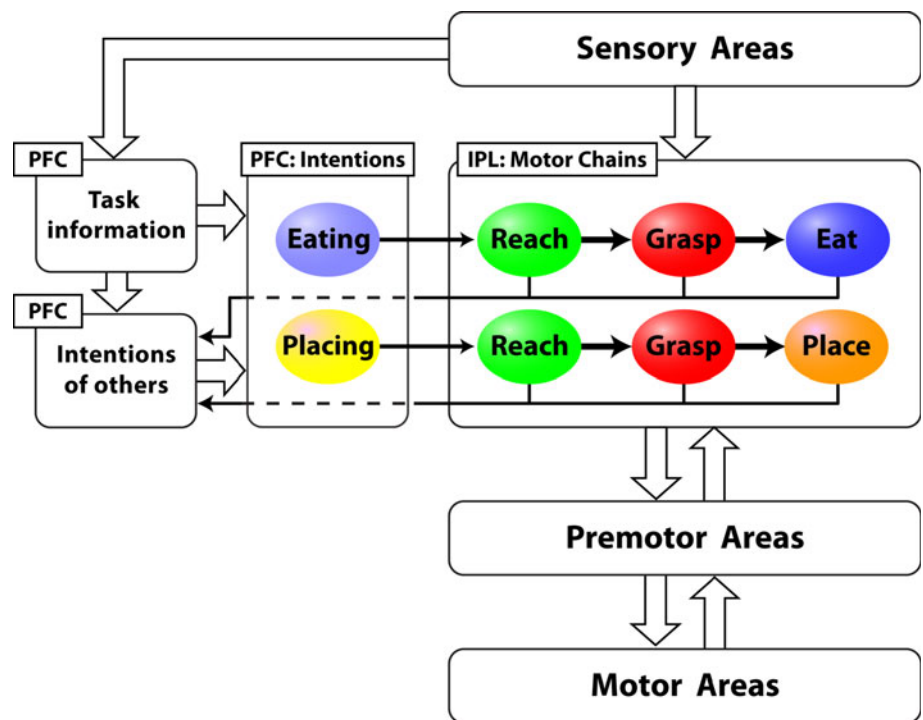


Fig. 4 Schematic representation of the Chain Model as implemented in the current paper. Each ellipse represents a subpopulation of neurons coding a specific motor act in IPL or an intention in PFC. Inputs from PFC select the appropriate chain, while sensory, proprioceptive, and motor feedback signals regulate the transmission of activity waves within the chains



parietal cortex layer in which, as the in the original model, action sequences are encoded in form of neuronal chains and a prefrontal cortex layer that contains ensembles of neurons that encode different types of information. On the one side, there are task-relevant information such as association rules, visual cues, and past events; on the other, there are (hypothetical) intentions of others. In addition to these, there are pools that encode own intentions. In the present model, both types of intention neurons receive inputs from other context neurons, which contribute to their preactivation before the actual movement begins. In this

sense, context information functions as prior or precursor for the formation of own and others' intentions. Considering this particular task, intentions of others can be viewed as a particular form of context, in the sense that according to the perceived intention or goal of the first individual, the second agent may produce one response instead of another.

Throughout the present work, *intention* is synonym of a *desire* or an *inclination toward* executing an action that leads to a specific goal or end state. It is important to note that having an intention does not lead automatically to the execution of the corresponding action as for this to happen,

specific physical conditions have to be satisfied. More specifically, the presence of an object may lead to the desire to grasp it, nevertheless to do so the object has to be not only visible but also at reaching distance. The mechanism that underlies the initiation movements is rather complex and involves several brain areas, such as the SMA (Luppino and Rizzolatti 2000), the BG (Hauber 1998) and the cerebellum (Thach 1975). For sake of simplicity, in the present model, this aspect will not be described in detail.

Finally, each element in the parietal layer, besides being part of a chain, receives input from sensory areas (in this scheme comprising STS) and sends output to lower-level motor areas (premotor and subsequently motor areas), which at the same time provide feedback signals about the ongoing action.

Mathematical details

Each unit in the different layers of the network represents a small subpopulation of neurons encoding specific motor acts (in IPL) or intentions and decisions (in PFC). The behavior of each neuronal pool is described by a firing rate model with detailed synaptic currents (Chersi et al. 2010; Dayan and Abbott 2001). This allows to compactly represent complex interactions between excitatory and inhibitory neurons within pools and to explicitly take into account the dynamics of ionic currents and neurotransmitters. The set of equations governing the behavior of a single neuronal pool is the following:

$$\begin{cases} \tau_v \frac{dv}{dt} = v + g(I_{syn} - I_{fra}) + \eta \\ \tau_I \frac{dI_{syn}}{dt} = -I_{syn} + \sum_h W_h \cdot v_h \\ \frac{dI_{fra}}{dt} = \alpha \cdot v - \beta \end{cases} \quad (1)$$

where v is the mean firing rate of a pool and $\tau_v = 60$ ms the corresponding time constant, $g(\cdot)$ is the current-to-firing rate ($I - f$) response function of a pool, η is an additional term that simulates spontaneous activity, I_{syn} is the total synaptic current, and τ_I the corresponding time constant ($\tau_I = 200$ ms except for the other’s intention pools where $\tau_I = 700$ ms), I_{fra} is the firing rate adaptation current due to neurotransmitter depletion, W_h is the connection strength from unit h to the current unit. Activity in sensory and temporal areas has not been explicitly simulated; instead, its output has been modeled as a bell-shaped activity peak lasting 300 ms. The firing rate adaptation has been modeled as a current I_{fra} , that increases with the firing of the neuron (through α) and hyperpolarizes the neurons in a pool slowing down their spiking. This current relaxes to

zero at a rate of β . In this implementation, $\alpha \simeq 3.8 \cdot 10^{-11}$ A and $\beta \simeq 9.8 \cdot 10^{-10}$ A/s.

The current-to-firing rate response function of the pools has been modeled as:

$$\begin{cases} g(I) = g_0 \cdot \tanh[\gamma(I - I_{thr})] & \text{for } I > I_{thr} \\ g(I) = 0 & \text{for } I \leq I_{thr} \end{cases} \quad (2)$$

where g_0 determines the maximum firing rate and γ the steepness of the response function, and I_{thr} is the firing threshold below which no response is present. In this implementation $g_0 = 150$ Hz, $\gamma = 1.5 \times 10^9$ A⁻¹ and $I_{thr} = 4 \times 10^{-10}$ A. All the parameters in this model have been chosen in order to reproduce as close as possible biological data.

The total input to a pool in a motor chain is given by:

$$I_{pool} = W_{prev} v_{prev} + W_{ext} v_{ext} \quad (3)$$

where W_{prev} defines the strength of the connection arriving from the previous pool or from an intention pool in case of the first element in the chain ($W_{prev} \simeq 1.1 \cdot 10^{-11}$ A/Hz), while W_{ext} represents the input arriving from STS, sensory or downstream motor areas ($W_{ext} \simeq 1.1 \times 10^{-13}$ A/Hz).

The inputs to pools that represent the other’s intention comprise higher-level visual areas (e.g., inferior temporal gyrus) that inform about the presence of specific cues and the parietal areas (areas PF-PFG), which provide information about ongoing activity:

$$I_{intention} = W_{cue} v_{cue} + W_{reach} v_{reach} + W_{grasp} v_{grasp} + W_{move} v_{move} \quad (4)$$

The single weights have been calculated by means of an automatic optimization procedure so that each of the four contributions produces an increase of 25 Hz in the firing rate of the intention pool ($3.9 \times 10^{-14} < W < 2.9 \times 10^{-12}$ A/Hz). The final effect is that as actions unfold neuronal pools in the prefrontal cortex that encode compatible goals/intentions increase their firing rate, while activity of incompatible pools decreases. In other words, the activity level of each pool represents the degree of similarity between the motor sequence encoded in the parietal cortex and the observed action. This activity level may be interpreted as an estimation of the intention or the goal of the observed individual.

Results

An important premise has to be made here. Action understanding is a complex phenomenon that can be analyzed at different levels of hierarchy and complexity. In particular for this task, the following distinctions have been made. There are two major categories: “simultaneous joint action” (transporting a package together) and “subsequent

joint action” (passing an object to someone who then finishes the action). In general, both can be “collaborative” (achieve something together) or “competitive” (steal something from someone) and can be “congruent” (push a closet together on the same side) or “complementary” (one pushes and the other pulls a table on opposite sides) (Sebanz et al. 2006).

This section will focus only on the simulation of “subsequent collaborative complementary joint actions” as described in the “Experimental setup” section. In a first condition, two different chains will be activated: one for the recognition and one for execution. Note that the second one may be a motor chain as well as a mirror chain (but not the same as the first one). In the second condition, the same chain will be activated both for the recognition and for the execution (thus providing an example of true mirroring behavior).

Condition 1: visuomotor activation

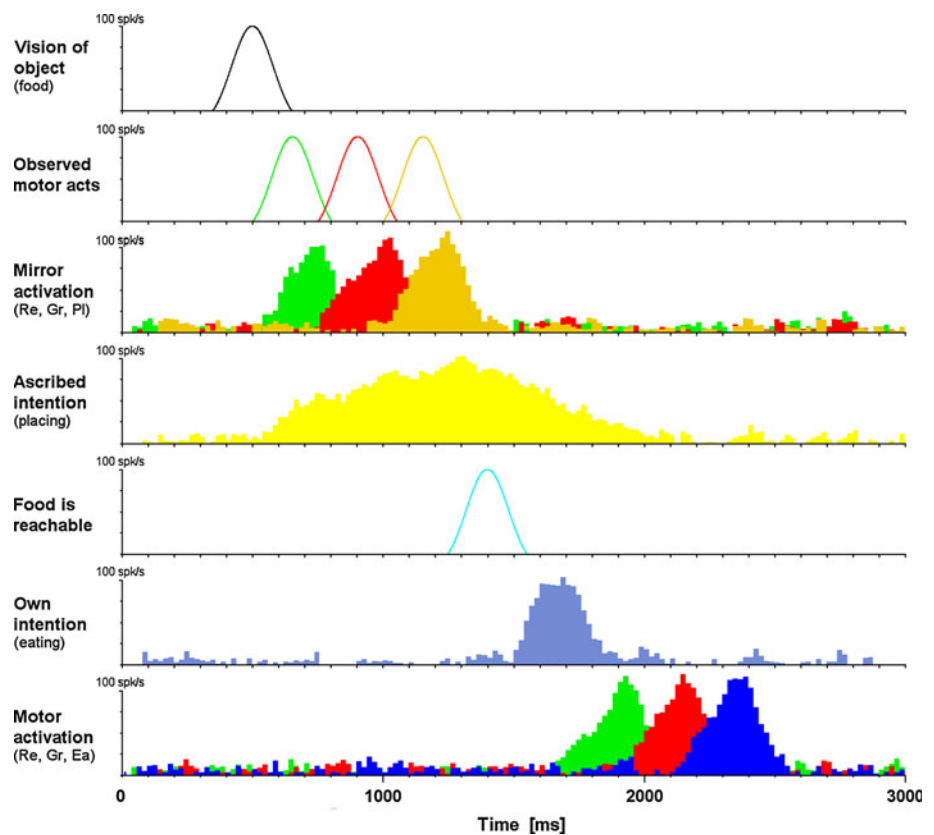
In the first condition, the first individual reaches and grasps a piece of food and hands it over to the other individual who grasps it and brings it to his mouth in order to eat it. From the point of view of the second individual (i.e., the observer), the key events in this sequence are the following: the vision of the piece of food, which provides a first

indication about the type of task and produces the preactivation of the “ascribed intention” pool, then the first individual begins to move his arm in order to reach and grasp the piece of food and hand it over. This produces in the observing individual the activation of specific pools in STS conveying to IPL information about the ongoing actions, which in turn increases the activation of the “ascribed intention” pool.

The results of the first phase are shown in Fig. 5. The top black bell-shaped curve represents the input from (simulated) higher-order visual areas informing about the detection of the cue (a piece of food). Without loss of generality, here, it is assumed that this signal is bell-shaped and transitory due to habituation or shift of attention focus. The green, red, and orange bell-shaped thin curves are the inputs from STS encoding the observed motor acts (green = reach, red = grasp, orange = place). The histograms below represent the response of the mirror pools to the visual input. Note that this is not a simple copy of the STS signal because each pool in order to be significantly active requires input from sensory areas as well as from the previous pool.

The yellow histogram in the middle represents the activity of the (estimated) “intention of the other” pool. Differently to the other pools in the network, the activity of these intention pools rises and decays slowly as it is the

Fig. 5 Firing rates as a function of time of neuronal pools and sensory input in the implemented network during the joint placing-eating task. Thin bell-shaped curves represent simulated sensory inputs (from IT and STS) that convey information about the cues and the ongoing action. Histograms represent the activity of intention, mirror, and motor pools (*green* reaching, *red* grasping, *orange* placing, *blue* bringing to the mouth, *yellow* intention to place, *light blue* intention to eat)



result of the integration of multiple neural signals and because the longer time constant that has been chosen for this type of pools in order to reproduce the long time span typical of working memory. It can be seen that the “estimated intention” pool reaches its maximal activity shortly after the visual input ends. In this simulation, the value of 100 Hz has been chosen as the threshold for the encoding of complete certainty about the intention of the other individual.

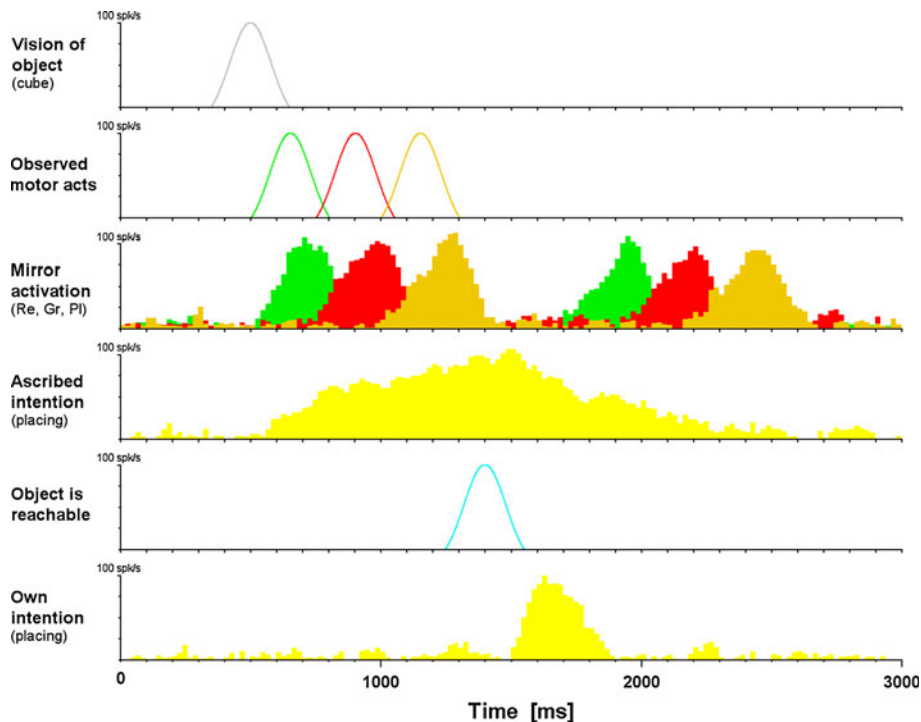
In the second phase, the observing individual has to produce a matching response to the action of the first individual: in this case, pick up the piece of food and bring it to his mouth. The response action sequence is already part of the agent’s motor repertoire, so only the choice of the correct motor chain to activate has to be made. The mechanism is straightforward: the combination of a specific context (the presence of food) and of the observed action (another individual placing the food at reachable distance) gives as a result the activation of the intention to execute the corresponding complementary action (taking the food and eating it). In the proposed network, it is assumed that the rule that associates these two preconditions to the selection of the corresponding motor sequence had been learned by the second agent before the start of the experiment. The results of this simulation are reported in the right part of Fig. 5 (approximately for $t > 1,500$ ms). The cyan bell-shaped peak represents the appearance of food at a reachable distance, while the light blue histogram shows the formation of the own intention. The green, red,

and blue histograms at the bottom represent the activation of the motor chain that encodes the whole response sequence. This histogram should be compared with the one in the right panel of Fig. 2. Note that in this condition, these elements receive input only from the previous pool and feedback from premotor areas.

Condition 2: (true) mirror activation

In the second condition, the first subject takes an object and places it near to the second individual, who takes it and places it into a container nearby. The neural activation is similar to the previous condition. This is exemplified in Fig. 6, which shows the firing rates as a function of time of all the neuronal pools in the brain of the second individual during the whole task. One important difference is that, in this case, there is a first activation of the mirror chain in “observation modality” leading to a placing action in the initial phase and the reactivation of the same mirror chain in “motor modality” leading again to a placing action in the second phase of the trial. In this condition, the elements of the mirror chain receive input from STS only during the observation phase, instead during the execution phase they receive only sensory and proprioceptive inputs. Important to note is the fact that the same chain is active during the two phases of the task but there is an overt motor output only in the second one. The assumption adopted in this model is to assign other brain areas (such as SMA or BG) the task of low-level motor initiation and inhibition,

Fig. 6 Behavior of the system during a joint action task that requires the generation of a response action executed by the same mirror chain active during observation. In comparison with the previous condition, the activation of the own intention to place the received object elicits the reactivation of the mirror chain (that encodes grasping to place) instead of the motor chain (that encodes a grasping to eat sequence)



leaving the IPL the role of motor sequence storage and retrieval.

Incomplete sequences

In order to verify the performance of the system when the observed motor sequence is incomplete or partially unrecognizable, the tests described in the previous section have been modified by providing the network with the same input sequence as before but without the last element (i.e., the “placing” motor act). The rationale behind this test is to verify whether the external input is essential for the successful propagation of activity along the chains and how incomplete input affects recognition and the formation of “ascribed intention”.

Results are shown in Fig. 7. As the second individual observes the action unfolding, the pool that represents the estimated intention of the first (acting) individual increases its firing rate. When the action terminates prematurely, the input from the second mirror pools activates only partially the third mirror pool, this leads the “estimated intention” pool to reach approximately a firing rate of only 80 Hz compared with the 100 Hz representing complete certainty. This value, besides indicating the confidence of the observer, can also be interpreted as a measure of similarity between the observed action and the sequence present in the motor repertoire.

Note that each simulation is based on a single observation. Since the system has no memory about past events, additional trials would not lead to an increase in performance.

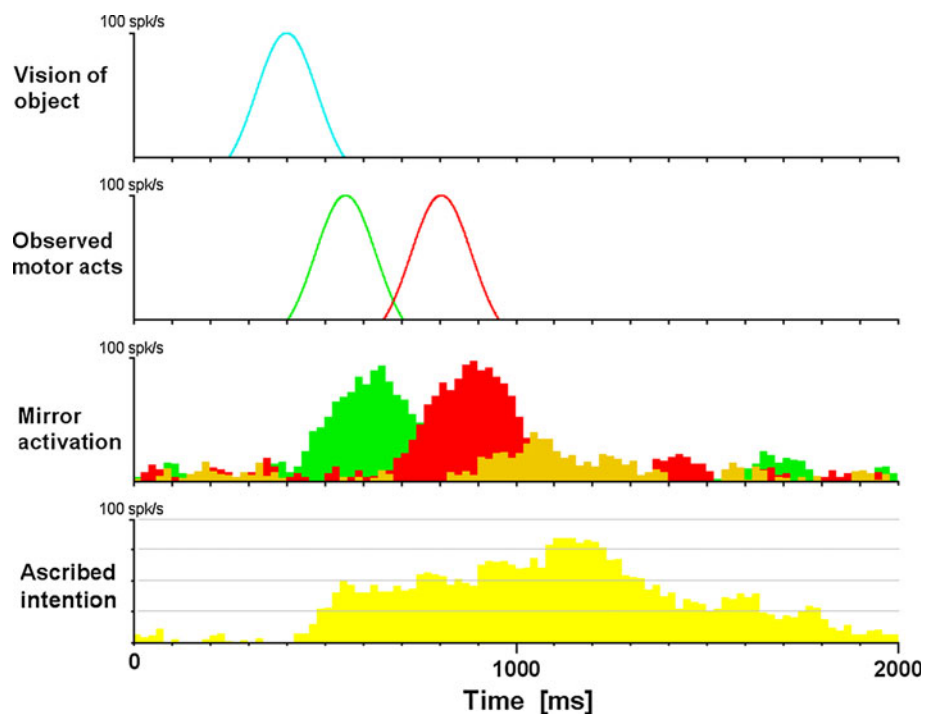
Discussion and conclusion

A fundamental requirement for successful joint action between individuals is the capacity to track and interpret the gestures of others and to smoothly and efficiently switch between observation and the corresponding execution phases.

Recent imaging and neurophysiological experiments have shown that, contrary to classical beliefs, the brain exploits the same circuit, the mirror neuron system, both for action execution and observation. This has two immediate fundamental advantages: the first one is the reduction in space and wiring, the second one is that this unifies the description of the motor aspects of the world thus eliminating the need for a continuous synchronization and “translation” from one representation to the other.

Building upon single neuron recordings of visuomotor neurons in the monkey parietal cortex, this paper provides a detailed description of how, in the framework of joint action, the mirror neuron system may handle goal inference and motor execution. In particular, it shows that a conceptually simple but biologically realistic neural network (the Chain Model) in which motor sequences are encoded by chains of neurons is easily capable of recognizing actions done by others and to produce an adequate motor response. Moreover, it proposes a computational framework capable of explaining how neuronal activity originating from the mirror circuit may be used in the prefrontal cortex to construct or evaluate hypotheses about the intentions of other acting individuals.

Fig. 7 Simulation of the response of the network to the observation of an incomplete sequence. The appearance of the cue and the first two observed motor acts produce an increase in the firing rate of the ascribed intention pool, which although, cannot reach the highest activity level because of the missing element of the sequence



On a more conceptual level, this work sustains the idea that because of the strong involvement of the MNS in low-level visual and motor understanding, its fundamental task may be to intervene predominantly in processes where agentive or meta-representational understanding is not required or available.

Among recent studies that have addressed a similar topic particularly noteworthy is the one of Oztop and colleagues (2005). This work describes a visuomotor architecture derived from classical control theory that uses a forward model to generate actions and infer the goal of observed ones. There are, though, some fundamental differences with the work presented here. The first one is that the former focuses on the visual recognition of hand trajectories during the reaching phase instead of on interpretation of entire action sequences composed of different motor acts. The second one is that it reduces the mental state inference process to the estimation of the intended target of a reaching movement. The third is that, as the authors state, it is a system level model lacking of a real biological instantiation. In contrast to this work, a common aspect of the above and other studies (Haruno et al. 2001; Demiris and Hayes 2002) is a general weakness in biological plausibility (mostly given by the fact that important neuroscientific discoveries have been made after their publication) and a very “engineeristic” approach to the matter.

The analysis of the functioning of this model and the interpretation of the outputs have given rise to important questions and to a variety of predictions that could guide future experimental research. In particular:

- The ability to successfully and smoothly join subsequent actions should increase with training and, in this model, would be due to the adjustment of the connection weights from the “context” and “others’ intention” pools to the “own intention” pool.
- The mechanisms described in the above sections are highly “cognitive” in the sense that they involve the active participation of prefrontal areas in monitoring the observed action, evaluating the intention of others and triggering the adequate response. It is still not clear if and to what extent animals are capable of such high level of mind reading. A possible alternative could be the establishment of a direct connection between the “recognition chain” and the “execution chain” (see Fig. 3), directly in the parietal cortex or through subcortical pathways, thus somehow bypassing the cognitive component. This would result in a classic action–reaction mechanism, which probably exists also in humans for habitual behaviors.
- Although in the simulated experiments, intention neurons in the PFC of the agents encode only medio-proximal goals (i.e., “placing” or “bringing to the

mouth”), most likely in human brains there exist, probably as an emergent process, neurons that encode the distal goal of the combined sequences (i.e., “eating” as a combination of the other’s “placing” plus my “bringing to the mouth”).

- Interference and facilitation phenomena may occur during the various phases of joint action when the same sequences, goals, or even effectors are involved, similarly to what happens during language processing (Chersi et al. 2010). To account for this, a firing rate adaptation component due to neurotransmitter depletion has been included in the equations (see Eq. 1).

Notwithstanding these interesting results, it is perfectly clear that the mechanisms coming into play during the observation of actions and even more during decision making are much more complex than depicted here. An important topic of current investigation is the development of a more complex architecture of PFC that allows the representation and processing of multiple and conflicting contextual and sensory information. Another important aspect that has been intentionally ignored here but will be tackled soon is how motor and mirror chains are built and how the decision rules are learned.

As a final note, an important aim of this article was also to stress the importance of computational models in neuroscience and cognitive sciences, as they may be very powerful tools for verifying the validity of hypotheses and for providing predictions for experimental research.

Acknowledgments A particular thank goes to Leonardo Fogassi and Pier Francesco Ferrari who gave me the opportunity to work in their laboratory at the University of Parma. This work was partially supported by the European project Artesimit, EC Grant IST-2000-29689.

Conflict of interest The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the sts region. *Trends Cogn Sci* 4:267–278
- Bonaiuto J, Rosta E, Arbib MA (2007) Extending the mirror neuron system model, I Audible actions and invisible grasps. *Biol Cybern* 96:9–38
- Bonini L, Rozzi S, Ugolotti F, Simone L, Ferrari PF, Fogassi L (2010) Ventral premotor and inferior parietal cortices make distinct contribution to action organization and intention understanding. *Cereb Cortex* 20:1372–1385
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fmri study. *Eur J Neurosci* 13:400–404
- Chalmeau R (1994) Do chimpanzees cooperate in a learning task? *Primates* 35:385–392

- Chersi F, Fogassi L, Rozzi S, Rizzolatti G, Ferrari P (2005) Neuronal chains for actions in the parietal lobe: a computational model. *Soc Neurosci Abs* 4128
- Chersi F, Mukovskiy A, Fogassi L, Ferrari PF, Erlhagen W (2006) A model of intention understanding based on learned chains of motor acts in the parietal lobe. *Comput Neurosci* 69:48
- Chersi F, Thill S, Ziemke T, Borghi AM (2010) Sentence processing: linking language to motor chains. *Front Neurobot* 4:1–12
- Craighero L, Metta G, Sandini G, Fadiga L (2007) The mirror-neurons system: data and models. *Prog Brain Res* 164:39–59
- Dayan P, Abbott LF (2001) *Theoretical neuroscience. Computational and mathematical modelling of neural systems*. MIT Press, Cambridge
- Demiris Y, Hayes G (2002) Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model. MIT Press, Cambridge
- Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. *Exp Brain Res* 91:176–180
- Fagg AH, Arbib MA (1998) Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw* 11:1277–1303
- Fogassi L, Bonini L, Simone L, Ugolotti F, Ruggeri E, Rozzi S, Chersi F, Rizzolatti G, Ferrari PF (2007) Time course of neuronal activity reflecting the final goal of observed and executed action sequences in monkey parietal and premotor cortex. *Soc Neurosci Abs* 636:6
- Fogassi L, Ferrari P, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005a) Parietal lobe: from action organization to intention understanding. *Science* 308:662–667
- Fogassi L, Ferrari P, Gesierich B, Rozzi S, Chersi F, Rizzolatti G (2005b) Supporting material for: Parietal lobe: from action organization to intention understanding. *Science* 308. <http://www.sciencemag.org/content/308/5722/662/suppl/DC1>
- Frith CD, Frith U (1999) Interacting minds—a biological basis. *Science* 286(5445):1692–1695
- Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. *Brain* 119:593–609
- Gallese V, Fadiga L, Fogassi L, Rizzolatti G (2002) Action representation and the inferior parietal lobule. In: Prinz W, Hommel B (eds) *Common mechanisms in perception and action: attention and performance*. Oxford University Press, Oxford. pp 334–355
- Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. *Trend Cogn Sci* 8:396–403
- Gentilucci M, Rizzolatti G (1990) Cortical motor control of arm and hand movements. In: Goodale MA (eds) *Vision and action: the control of grasping*. Ablex, Norwood. pp 147–162
- Grinnel J, Packer C, Pusey AE (1995) Cooperation in male lions: kinship, reciprocity, or mutualism? *Animal Behav* 49:95–105
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000) Brain areas involved in perception of biological motion. *J Cogn Neurosci* 12:711–720
- Haruno M, Wolpert DM, Kawato M (2001) Mosaic model for sensorimotor learning and control. *Neural Comput* 13:2201–2220
- Hauber W (1998) Involvement of basal ganglia transmitter systems in movement initiation. *Prog Neurobiol* 45:507–540
- Iacoboni M, Koski LM, Brass M, Bekkering H, Woods RP, Dubeau MC, Mazziotta JC, Rizzolatti G (2001) Reafferent copies of imitated actions in the right superior temporal cortex. *Proc Natl Acad Sci USA* 98(24):13995–13999
- Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G (2005) Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol* 3(3):79
- Luppino G, Rizzolatti G (2000) The organization of the frontal motor cortex. *News Physiol Sci* 15:219–224
- Melis AP, Hare B, Tomasello M (2006) Chimpanzees recruit the best collaborators. *Science* 311:1297–1300
- Nishitani N, Hari R (2000) Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci USA* 97(2):913–918
- Noe R (2006) Cooperation experiments: coordination through communication versus acting apart together. *Animal Behav* 71:1–18
- Oram MW, Perrett DI (1994) Responses of anterior superior temporal polysensory (stpa) neurons to biological motion stimuli. *J Cogn Neurosci* 6:99–116
- Oztop E, Arbib MA (2002) Schema design and implementation of the grasp-related mirror neuron system. *Biol Cybern* 87(2):116–140
- Oztop E, Wolpert DM, Kawato MM (2005) Mental state inference using visual control parameters. *Cogn Brain Res* 22(2):129–151
- Perrett D, Harries M, Bevan R, Thomas S, Benson P, Mistlin A, Chitty A, Hietanen J, Ortega J (1989) Frameworks of analysis for the neural representation of animate objects and actions. *J Exp Biol* 146:87–113
- Rizzolatti G, Craighero L (2004) The mirror neuron system. *Annual Rev Neurosci* 27:169–192
- Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2:661–670
- Rushworth MFS, Behrens TEJ, Johansen-Berg H (2006) Connection patterns distinguish three regions of human parietal cortex. *Cereb Cortex* 16:1418–1430
- Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. *Trend Cogn Sci* 10(2):70–76
- Sebanz N, Knoblich G (2009) Prediction in joint action: What, when, and where. *Top Cogn Sci* 1:353–367
- Thach WT (1975) Timing of activity in cerebellar dentate nucleus and cerebral motor cortex during prompt volitional movement. *Brain Res* 88:233–241
- Visalberghi E, Quarantotti BP, Tranchida F (2000) Solving a cooperation task without taking into account the partner's behavior: the case of capuchin monkeys (*cebus apella*). *J Comp Psychol* 114:297–301