Neuroscience Letters 540 (2013) 43–55

Contents lists available at SciVerse ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Mini review

Mirror neurons: Functions, mechanisms and models

Erhan Oztop[a,b,*], Mitsuo Kawato[b], Michael A. Arbib[c]

[a] Ozyegin University, Istanbul, Turkey
[b] Advanced Telecommunications Research Institute International, Kyoto, Japan
[c] University of Southern California, Los Angeles, CA 90089, USA

HIGHLIGHTS

- In the literature various *functions* and supporting *mechanisms* are attributed to MNs.
- Many of the *functions* are not observed in monkeys: look for evolutionary explanation.
- The distinction between a brain function and mechanism must be made clear.
- Computational models can be used in clarifying mechanisms that can support MNs.

ARTICLE INFO

Article history:
Received 7 May 2012
Received in revised form
27 September 2012
Accepted 2 October 2012

Keywords:
Mirror neuron
Computational model
Action recognition, imitation, language evolution
Mirror neuron development
Direct matching

ABSTRACT

Mirror neurons for manipulation fire both when the animal manipulates an object in a specific way and when it sees another animal (or the experimenter) perform an action that is more or less similar. Such neurons were originally found in macaque monkeys, in the ventral premotor cortex, area F5 and later also in the inferior parietal lobule. Recent neuroimaging data indicate that the adult human brain is endowed with a "mirror neuron system," putatively containing mirror neurons and other neurons, for matching the observation and execution of actions. Mirror neurons may serve action recognition in monkeys as well as humans, whereas their putative role in imitation and language may be realized in human but not in monkey. This article shows the important role of computational models in providing sufficient and causal explanations for the observed phenomena involving mirror systems and the learning processes which form them, and underlines the need for additional circuitry to lift up the monkey mirror neuron circuit to sustain the posited cognitive functions attributed to the human mirror neuron system.

1. Introduction

Mirror neurons for manipulation fire both when the animal manipulates an object in a specific way and when it sees another animal (or the experimenter) perform an action that is more or less similar. Such neurons were originally found in macaque monkeys, in the ventral premotor cortex, area F5 [21,32,76], and later also in the inferior parietal lobule [28,78]. Not all visuomotor neurons in F5 show the mirror property. There are F5 visuomotor neurons that selectively discharge to the visual presentation of a given object, which also discharge selectively during grasping of that object [61]. These neurons are called canonical neurons, and are believed to play a crucial role in transforming visual appearance of objects into motor plans for interacting with them [77]. Area F5 also includes auditory mirror neurons [54] that respond not only to the view but also to the sound of actions with typical sounds (e.g. breaking a peanut, tearing paper). Although classically, the actions associated with mirror neurons in the monkey are considered to be transitive, i.e. the action has to be directed to an object, which may be removed from the view of the monkey before hand-contact [e.g. 95], more recent data indicates that transitivity may not be a prerequisite for mirror-like activity (see also Section 7).

Recent neuroimaging findings indicate that the adult human brain is endowed with a system for matching the observation and execution of actions which might be homologous to the macaque mirror neuron system [10,73,77]. In spite of the growing number of human brain imaging data related to posited 'human mirror systems' [e.g. 6,13,25,43,44,83], the experimental data on mirror *neurons* are available mainly for monkeys as systematic recordings using electrophysiology cannot be used to investigate the human brain. Therefore it must be emphasized that in humans a *mirror system* refers to a brain region (or set of brain regions) that becomes active for both observation and execution of a class of actions.

In the literature a set of *functions* is attributed to monkey and human putative mirror neuron system, and several terms are used to describe *mechanisms* underlying these functions. However, many

* Corresponding author at: Ozyegin University, Faculty of Engineering, Nisantepe Mah., Orman Sk. No. 13, Alemdag-Cekmekoy, Istanbul, Turkey.
Tel.: +90 216 5649392; fax: +90 216 5649057.
E-mail address: erhan.oztop@ozyegin.edu.tr (E. Oztop).

0304-3940/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neulet.2012.10.005

Box 1: The distinction between function and mechanisms becomes apparent when seen as answers to key scientific questions

Function of MNs: what MNs are used for by the central nervous system? What is the purpose, use, role in the brain?
Mechanism yielding MNs: Why MNs exhibit the properties we observe? What is the causal rather than teleological or essentialist explanation of MN properties?

of these functions are observed in human but not in monkeys, thus suggesting evolution within the mirror systems or within the wider networks of which they are part. Among these are imitation [e.g. 13,58], action understanding [e.g. 95], intention attribution [43] and (evolution of) language [74]. Recently, reviews and meta-analyses that are critical of the claimed mirror neuron functions have started to appear, in particular with the focus on the ambiguity of the terminology used to describe mirror neuron functions such as direct matching and motor resonance [17,19,23,92]. In part such failures to unambiguously describe mirror neuron function, follow from ignoring the distinction between a *brain function* and a possible *mechanism* (Box 1). In the majority of mirror neuron literature, functions associated with a mirror system in humans are attributed to "direct matching" or "motor resonance" and sometimes with "motor simulation" as a mechanism to underlie action/intention understanding [34,78] and theory of mind [33] without either a precise definition of such a "mechanism" nor a clear account of how it contributes to the observed function. It is simply assumed that mirror systems are involved in this property. However, we know (by extrapolation from the macaque) that a mirror system will contain many types of neurons other than mirror neurons. Thus, when a brain imaging study reports increased activity in a mirror system for some task relative to a control it is a mistake (all too frequent in the literature) to assume that activity in mirror neurons underlies the activation – in some cases, it will be, but by no means in all.

Our task in this article is to make the case for the use of computational models, whether in terms of neural networks or higher level constructs such as control systems, in clarifying mechanisms that are sufficient to explain observed phenomena involving mirror systems. The mechanisms we demonstrate might not be the same with those employed by the brain, but their very precision sets the stage for experiments more precise than those guided by claims like "a motor resonance supports the observed function." To be concrete, we will be interested with these questions:

- Is the postulated mechanism sound from a computational point of view?
- To what extent can a brain implement the postulated mechanism?
- What are the additional mechanisms needed on top of the claimed mechanisms to yield the posited functions?

2. Direct matching

To evaluate the direct matching or motor resonance hypothesis, we must first clarify *what is matched (decoded)*? and *how is the matched action encoded* by the mirror neurons (MNs).

2.1. What is decoded by MNs?

When discussing the decoding used by MNs, the main distinction is whether the code is a motor code or not. The secondary concern is the level of detail it entails. Here we propose three types of encoding that is often used in the literature (Box 2). L1 and L2 refer to motor encodings where the former involves a detailed motor parameter set. L3 refers to an abstract representation that may be used to generate action but by itself does not constitute a motor plan.

Box 2: Main interpretations as to what is decoded by the MNs in the literature

L1: Decoding the detailed motor parameters of an observed act (e.g. the trajectory of the hand)
L2: Decoding the schema level motor plan (e.g. put object in cup)
L3: Decoding the intention (goal of the action) (e.g. wants to eat the peanut)

Proponents of an imitation role for mirror neurons would adopt the view that direct matching allows the automatic generation of the motor parameters of an observed action (view L1). However activation of MNs alone does not induce movement in monkeys, and during execution the canonical neurons (a subset of F5 neurons that are not mirror neurons) become active in addition to mirror neurons. So some difference in neural circuitry between monkey and human would be required to activate canonical neurons during imitation to turn observation into action. The emulation (i.e. goal level imitation) role proponents would go with L2 where decoding is limited to a rough plan of the observed action. Finally, for those that assign the understanding or intention inference role to mirror neurons, view L3 would be the most suitable explanation.

Direct matching views of L1 and L2 must be considered strictly different from that of goal or intention inference [e.g. 28,43] view of L3; because, in the former views what is decoded is a motor code that will more or less regenerate the observed act, and this may be very different from that of the action needed to generate the intended effect of the actor (decoded by the MNs according to L3). For example, the actor may only have the means of pushing a lever to tip over a vase while the observer has only the means of pulling the lever to tip it over. The motor codes of the observer and the actor have to generate opposite actions for the same intention of tipping over the vase. One can circumvent the conflict between L3 and other views by adopting a dual route imitation mechanisms (see Fig. 1) in which the given example will engage the ventral (i.e. not involving the mirror neurons) imitation pathway. Hickok and Poeppel [39,40] distinguish a dorsal and ventral route in speech perception, where the former is linked to the phonemic structure of what is heard while the other relates to the phonemes in the word – corresponding, perhaps to different imitation goals of (ventral) repeating the word you have heard versus (dorsal) repeating a nonword or imitating an accent.

This feeds into the general issue of whether the mirror system is necessary for repeating an action or simply enriches perception of an action when it is in the observer's motor repertoire. Here we do not claim an imitation role for mirror neurons but if they are indeed involved in imitation of novel actions, then the notion

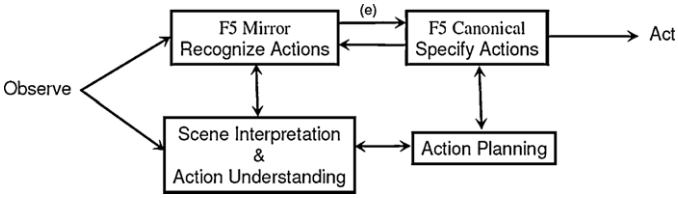

Fig. 1. The perceptuomotor coding for both observation and execution contained in F5 region for manual actions in the monkey is linked to "conceptual systems" for understanding and planning of such actions within the broader framework of scene interpretation. The interpretation and planning systems themselves do not have the mirror property save through their linkage to the actual mirror system.

of mirror neurons must be expanded to include actions that are not in the observer's repertoire. This however may not reflect the reality as when the observed actions are not possible for the observing human, it appears that mentalizing mechanisms, relying neural structures beyond mirror systems, are additionally recruited to process the actions [6].

Umilta et al. [94] trained macaque monkeys to use inverse pliers for picking up peanuts (i.e. squeezing the pliers opened the jaws of the pliers), and found that mirror neuron activity does not correlate with the hand movements involved in grasping the tool but rather with the motion of tips of the pliers. This has been interpreted as decoding the goal of the action (i.e. grasping a peanut) which is compatible with the intention coding view of parietal mirror neurons [28]. More generally, Arbib et al. [4] links tool use to the distalization of the end-effector, which brings trajectory (not just overall goal) back into play, and relates their analysis both to a comparable case of human tool use [46] as well as to a range of computational models for generation and recognition of hand actions.

In their meta analysis of action observation and execution, de Vignemont and Haggard [19] start off by assuming that there *is* a shared representation of an action between the executor and the observer and reach the conclusion that L2 type of view is best explanatory for the mirror neuron data reported in the literature. This would be incompatible with the intention/goal decoding views of Fogassi et al. [28], and Umilta et al. [94] since the decoded intention/goal may involve different and even incongruent motor codes. Note however that theirs is an unduly strong assumption. The truth of the *existence of a sharing* relies on the nature of coding for execution and observation. These two conditions never generate the same neural activity, even for strictly congruent mirror neurons, which is often not spelled out but can be directly seen from the raster plots representing the mirror neuron firing.

Summing up, it would not be wrong to say that L1 (low level motor parameter coding) view is out of fashion and the trend is moving towards L3 (intention coding) among neuroscientists, for which we have more to say in Section 8. Note that the latter is different from what Goldman and Gallese [33] proposed for understanding intentions. There, the MNs were taken to be either L1 or L2 type and used only to decode the motor related information (m_{actor}) from the observed action. The intention/goal assignment was due to the pretend decision making based on a candidate intention which outputs an action code ($m_{imagery}$). The match between these two motor representations then indicates that the demonstrator's intention is the one that has been fed into the decision making system. Gallese and Goldman [33] states that mirror neuron activity resembles the simulation heuristic in that the observer obtains a state that matches that of the observed. So, 'simulation' here does not refer to the minute-to-minute mentally executing a movement reminiscent of mental rotation tasks. However, it is not clear whether the mirroring function or its use in the process of decision making is referred as simulation (see Section 6 for more on this).

It would be misleading to close the case as to what mirror neurons encode based on the most recent and popular view; instead a computational thinking must be adopted to interpret the data. The concept of 'population vector' has been very useful in explaining how the reach direction may be encoded accurately with the broadly tuned firing rate of a set of neurons in the primary motor cortex [35]. The range of congruency observed in mirror responses could be explained by such a distributed representation for the features of an observed action, where each mirror neuron has a preferred action feature, analogous to the preferred direction of Georgopoulos [35]. MN research desperately needs such population level analyses to uncover which features of an observed action are encoded in the MN activity.

2.2. Inverse modeling

The central nervous system uses internal models for movement planning, control, and learning [50,98]. A forward model is one that predicts the sensory consequences of a motor command [51,97,99]; while an inverse model transforms a desired sensory state into a motor command that can achieve it. From a computational point, the direct matching hypothesis is equivalent to saying that MNs perform internal inverse modeling [49], where the MN code can be used for action generation. This action will be a faithful replication, including the intricate details for the L1 version of the hypothesis. For the L2 version, the action will be more or less the same, and utilize the same limbs; but, may have variation in its details (e.g. the way of grasping). Finally for the L3 version of the hypothesis, the action can be very different as long as the final goal, i.e. the intention coded, is satisfied. This of course, leaves much freedom which the execution mechanism has to fill in (e.g. decision to use the foot or the hand). Note that even in the case L1, sole MN activity is neither sufficient nor necessary to generate action, as the reversible lesion study of Fogassi et al. [29] shows. Inverse model learning at dynamics level is possible [37,98] and such mechanism may be at work yielding MN type activations [68]. Unfortunately, up to date there is no data showing that MN code can be used to generate an observed act. In fact, there is no hard evidence showing any behavioral effect of MN activation.

In thinking through this discussion, it is crucial to note the difference among primates in imitation capability [see 3, Chapter 7, for a recent review]:

- Monkeys have limited capacity for imitation perhaps facilitated by stimulus or receptor enhancement: the observed action draws the monkey's attention to a particular object or the use of a particular effector.
- Apes exhibit imitation by behavior parsing [11] (simple imitation): They may come to recognize the subgoals of another's behavior through repeated observation but do not pay attention to movements used by the other to achieve those goals.
- Humans can exhibit action level imitation (complex imitation), which combines subgoal recognition with copying the details of the manual actions another individual uses.

It is worthwhile investigating whether the levels of imitation ability across primates are due to what MNs extract with regards to L1–L3 coding, or the ability of the central nervous system to employ MN activity to guide imitation that exploits these features. In any case, recall the data of [94] in which mirror neurons were observed to encode the motion of the end effectors in grasping an object whether the grasp was conducted using normal pliers, reverse pliers or the bare hand – rather than the hand motion (or, a fortiori, the control of the hand) required to achieve that goal. However, they found a mixture of neurons in primary motor cortex when the monkey executed these three versions of the task – those correlated with the hand movement (opening or closing the hand) as well as those correlated with the motion of the end effector. This makes clear that a full computational/control-theoretic account linking inverse models to mirror neurons must extend to include a range of other circuitry.

2.3. Hebbian-like associative learning

A simple computational mechanism that is compatible with the L2 [L3] view of the hypothesis is a memory system where each memory item is composed of the motor plan [intention] and the resulting sensory stimulus. Then MNs can be conceptualized as retrieving the desired memory item either based on the sensory stimulus (e.g. visual action observation) or the

motor plan [intention] (action execution). Here, by adopting L2 or L3, we more or less define the nature of the motor representation that serves as input to MNs. For, however, the visual information arriving at MNs we have not specified any representation. In the literature, it is often and implicitly assumed that MNs have access to highly processed visual information through Superior Temporal Sulcus for body parts and biological motion [12,70,71] and Anterior Intraparietal area for the affordances of objects [80,81,88]. If we assume such a compact visual representation arriving at MNs then, such a memory system can be realized via associative learning implemented by a neural circuit. This can be described as Hebbian learning (i.e. strengthening the synaptic connection between two neurons that fire at approximately the same time) at a higher level. Keysers and Perrett [52] proposed a conceptual model based on this Hebbian association view, where the Superior Temporal Sulcus (STS) 'sees' the action, whereas the ventral premotor cortex (F5) 'executes' the action. The time proximity of the self-executed actions wire F5 and STS linking the motor code of the action with its sensory code (appearance). Hence, after adequate experience, whenever STS responds to the observation of a particular action, say X, F5 neurons that encode action X is activated hence the mirroring property for action X. A clear distinction between simple Hebbian association (that relies on time contiguity) of Keysers and Perrett [52] and a predictive association mechanism (that relies on contingency as well as time contiguity) is given by Heyes [38] and others (see Cooper, Cook, Dickinson, Heyes, this issue). We also think that the Hebbian strengthening mechanism must be gated by additional circuits to ensure that the associations formed are due to the sensorimotor dynamics generated by the organism in its environment. With this mind set, a neural network compatible with the L2 coding view was implemented enabling the visual appearances of the self generated hand postures of a robotic hand to be associated with the motor code generating them as a model of mimicry [16]. In fact similar ideas have been in use in robotics for some time [e.g. 7,56]. Note however that, the association learning may not be sufficient to sustain a mechanism to support L1 coding due to the sheer number of action possibilities (consider the continuum of the action parameters).

Above discussion indicates the computational feasibility of associative learning for yielding MN like neural units once a compact representation is assumed. However, these models are based on *event-level* association: the entire visually observed trajectory becomes paired with activation of mirror neurons encoding the action involved. Macaque mirror neurons, on the other hand, will in general respond earlier in the trajectory to the extent that the prefix of the observed trajectory is unambiguous. *Trajectory-level* associative learning is also possible (see for [68] examples) but is formidably difficult to scale up to serve as a model for human action decoding. Furthermore, although association learning view sounds plausible as a mechanism, it does not offer any hints as to the function of the mirror neurons in the macaque monkeys and other primates.

2.4. What is the coding used by MNs?

Stating that mirror neurons directly match observed acts into motor representation is awfully ambiguous. There are at least four candidates for how to read off a directly matched action:

T1: A single neuron's average activity
T2: Population level average activity
T3: Temporal pattern of single neuron activity
T4: Temporal pattern of population level activity

The temporal pattern of mirror neuron activity (T3 and T4 views) is not developed in detail in the experiments reported in the literature so far, though some of the figures allow certain inferences (and note the explicit appeal to population codes based on data on differential timing of activity of F5 neurons in the development of the FARS model [26] of visually directed grasping). Usually a neuron would be considered encoding an action or an intention as indicated by its averaged activity in the neighborhood of a relevant event marker, e.g. contact with the object [21,32,76]. This view, though helpful in the early investigations must be replaced by more advanced investigation techniques. Although the temporal aspect of MN activity still does not receive the much needed attention, the necessity of population level analysis is obvious; otherwise it would be difficult to accept brain imaging data since single neuron activity would not be picked by noninvasive imaging techniques [22]. So it can be inferred that current trend in interpreting MN activity is based on the T1–T2 views, but that the reality is most likely to match T4, and new neurophysiological data are needed whose design explores this perspective.

2.5. Temporal activity

The associative learning account of MNs reviewed in the previous section sounds like a good explanation except that one would not expect a premotor area, F5, to be the host of a visuomotor association center with no motor function. One natural, but often neglected hypothesis is that F5 mirror neurons are there for motor control [65,69]. This has received little attention from neuroscientists, probably due to the meticulous analysis of macaque data required to pin down (or reject) this function. This hypothesis implies that temporal patterns have critical information when analyzed together with the observed and executed movement. One may have the impression that the firing pattern of a mirror neuron that 'encodes' an action is the same when the action is executed or observed. This is not true. This is not true even for strictly congruent mirror neurons. Furthermore, this difference appears to persist even when the population average of firing activities is considered [28, Fig. 5].

3. Action understanding and direct matching

Mirror neurons, when initially discovered in macaques, were thought to be involved in action recognition [30,32,76]. Although the term "action understanding" was often used, the exact meaning of "understanding" as used is not clear. In fact, the neurophysiological data simply show that a mirror neuron fires both when the monkey executes a certain action and when he observes more or less congruent actions. We think at minimum "understanding" includes the ability of an organism to incorporate an external event into his future behavior plan for improving his chances for satiety, safety and fitness.

While it is true that mirror neuron activity correlates with *observing* an action, we suggest that such activation is insufficient for *understanding* the movement – thus the indication of other systems for interpretation and planning in Fig. 1. A possible analogy might be to observing a bodily gesture in a foreign culture – one might be able to recognize much of the related movements of head, body, arms and hands that constitute it, yet be unable to understand what it means within the culture. The Figure emphasizes that F5 (and presumably any human homologue labeled as a "mirror system") contains non-mirror neurons (here the canonical neurons are shown explicitly) but that it functions only within a broader context provided by other brain regions for understanding and planning of actions within a framework of interpretation of the current environmental and motivational context. The direct pathway (e) from mirror neurons to canonical neurons for the same action may yield "mirroring" of the observed action, but is normally

under inhibitory control. In some social circumstances, a certain amount of mirroring is appropriate, but the total lack of inhibition exhibited in *echopraxia* and *echolalia* [79] – the compulsive repetition of observed actions or heard words which in humans may accompany autism – is pathological.

Unfortunately, in almost all mirror neuron experiments, the monkey is not given the opportunity to show by his behavior that he understands the action observed. Without giving an explicit definition of action understanding, Rizzolatti and Craighero [75] suggests that mirror neurons mediate understanding of actions of others through a certain mechanism which they describe as: "*Each time an individual sees an action done by another individual, neurons that represent that action are activated in the observer's premotor cortex. This automatically induced, motor representation of the observed action corresponds to that which is spontaneously generated during active action and whose outcome is known to the acting individual. Thus, the mirror system transforms visual information into knowledge*". In a recent report Kilner et al. [53] proposed a Bayesian predictive role for mirror neurons that can undertake the action understanding role. However, they did not implement an actual computational model. We will see later that [69] earlier developed and implemented a computational model, which assigns a sensory forward predictor role to mirror neurons that supports action understanding.

For plausibility of the MN involvement in action understanding, it is critical to find out how the (ventral premotor) mirror activations would reach higher cognitive centers. One possibility is that reciprocal connections with parietal regions may carry this information to the prefrontal area. Then this will bring the possibility that parietal regions are responsible for action understanding, F5 being subordinate to it. The recently found parietal mirror neurons [31] may be involved in this function. This also fits well with the recent finding that parietal mirror neurons encode actions in a context dependent way suggesting that they encode intentions of the demonstrator [28].

We have seen that mirror neurons would respond to a grasping action even the last part of the movement is hidden from the monkey, provided that the monkey has seen that there is an object behind the occluding curtain (condition B) [94]. If there is no object behind the curtain (condition D) then there is no mirror neuron response. In their interpretation of this result, Rizzolatti and Craighero [75] state "*Note that from a physical point of view B and D are identical. It was therefore the understanding of the meaning of the observed actions that determined the discharge in the hidden condition*". But, then this says that *understanding* preceded the MN activity! (We do not claim this but see Csibra [17]). In the same publication they conclude "*... both the experiments showed that the activity of mirror neurons correlates with action understanding. The visual features of the observed actions are fundamental to trigger mirror neurons only insomuch as they allow the understanding of the observed actions. If action comprehension is possible on another basis* (e.g., *action sound), mirror neurons signal the action, even in the absence of visual stimuli*." If indeed the monkey understands the action then we agree that the MN activity is correlated with this understanding, as it decodes (to some extent) the observed action; but, there is no data indicating whether the understanding precedes MN activity or is caused by MN activity (also note that there is no hard proof that understanding ever takes place). Indeed, in modeling the Umilta et al.'s [95] result Bonaiuto et al. [9], make clear that the crucial distinction rests on the encoding within a working memory circuit, rather than in the mirror neurons, of whether or not an object was recently observed and has not been seen to be removed. The anatomical connections of F5 is compatible with this view; there are projections from area 46, a site of working memory, onto part of F5 where mirror neurons are located [36]. Here we diverge from Csibra [17] who suggests that the data indicates that the understanding process take place elsewhere and reported by the MNs. We simply hold that understanding cannot be mediated by MNs alone. If indeed MNs are involved in action understanding then we may say that they are part of an understanding network.

3.1. "Understanding" must affect behavior; otherwise it has no use for the organism

We hold that for understanding another's action, the MN activity must have the potential to change the monkey's future responses; otherwise it has no function at all. This change can be as simple as saccade or approach behavior associated with the recognition provided by the MNs. In the other extreme, it can be a deliberate planning and action based on the inferred internal state of the observed actor that is compatible with the action decoded by the MNs of the observing monkey. As in Fig. 1, both cases definitely require additional circuitry that MNs must project to transmit the motor code belonging to the observed action. This additional circuit would then perform the understanding related activities. If understanding role of mirror neurons is true, one would like to know how the MNs are wired within this larger understanding network.

A recent study investigated the connections of the macaque ventral premotor cortex, area F5 [36]. The study showed that the anterior sector of F5 (F5a) has robust prefrontal connections with the ventrolateral prefrontal cortex which is thought to be involved in high level action planning (area 46v), and non-spatial high level processing (area 12). The remaining part of F5 (posterior part of the postarcuate bank, F5p, and postarcuate convexity cortex, F5c) have weak connections to these areas. Current data indicates that area F5a is reciprocally connected with the prefrontal areas. However, the data does not help us to determine whether F5c (where mirror neurons are located) projects to these areas, since the tracers injected to F5c were retrograde (FB and CTBr) and retro-anterograde (LYD). It is possible that the (weak) dyeing obtained in the prefrontal areas 46 v and 12 is due to the afferents emanating from there with no or little output projecting to the prefrontal areas. In general, prefrontal → F5 connections appear to be stronger than F5 → prefrontal connections, of which F5a has the largest volume of connection with the dominant role of integrating sensory input and communicating with the premotor cortex [36,89]. Although, F5c seems to not contribute much to the prefrontal projections, it is worth to note that there are dense intra F5 connections [36] via which F5c can relay its decoded motor code to the prefrontal areas (e.g. via F5a). However, this is somewhat in conflict with the fact that F5a does not show MN-like responses.

4. But MNs are active during execution of the monkey's own actions

Our conclusion from the above discussion is that it is not meaningful to focus on MNs in isolation but their (possible) role in an understanding network must be the target for research, for which the current anatomical and behavioral data falls short of giving a full picture. However, even this broader discussion relates – as does almost all the extant literature on mirror neurons – to the activity of MNs during observation of another's actions, what about the activity of MNs during the execution of an intended action? Presumably, "understanding" was elaborated during the planning stage; but the ensuing activity during execution must contribute to something other than understanding. One of the very few attempts to answer this question – other than the MSI model we discuss in detail in a later section – introduced the novel hypothesis that a mirror system may contribute both to monitoring the success of a self-action and to recognition of one's own *apparent* actions if they deviate from one's intended actions [8]. The gist of the model, called augmented

competitive queuing (ACQ), is to consider action choice as based on the desirability of executable actions. The idea is this: When we start to execute an intended action within a certain context, mirror neurons can create an expectation of reaching the goal of that action. If the expectation is not satisfied, then the brain can decrease its estimate of the action's executability – of how likely it is to succeed in the given context. But if we fail to execute one action, we may nonetheless, in some cases, succeed in completing a movement and achieving a desirable goal (or taking a step towards such a goal). If so, the mirror system may "recognize" that the action looks like an action already in the repertoire. As a result, learning processes can increase the neural estimate of the desirability of carrying out that action when the animal attempts to achieve the goal in the given context. By expressing these ideas in a form that could be simulated on a computer, the modeling showed how this "what did I just do?" function of mirror neurons can contribute to the learning of both executability and desirability, and how in certain cases this can support rapid reorganization of motor programs in the face of disruptions.

5. Simulation theory and direct matching

We return to the emphasis on interpreting MNs in the context of observing the actions of others. Gallese and Goldman [33] suggest that the purpose of MNs is to enable an organism to detect certain mental states of observed conspecifics via mental simulation. According to this view, mirror neurons could be the precursor of mind-reading ability, being compatible with the simulation theory hypothesis according to which mental states of others are represented by representing their states in terms of one's own – this is in contrast to the "theory theory" which asserts that mental states are represented as inferred conjectures of the observer's naive theory of mental states.

According to Gallese and Goldman, the MNs are considered as intrinsic to a mental simulation routine, which work as follows (see Fig. 2). MNs activity (M_{actor}) represents the action demonstrated by an agent in motor terms (L1 or L2 of Section 2). The observer then generates a guess as to the (virtual) mental state or goal (g) which generated the observed activity, which he then feeds to his decision system. The output of the decision system is a hypothetical motor plan (M_{img}) which can be compared to M_{actor}. Then the observer can, by changing his initial guess, reduce the mismatch between M_{actor} and M_{img}, converging on to a mental state or a goal that achieves $M_{actor} \sim M_{img}$. In this conception, the comparison takes place in the motor domain, i.e. the output of mirror neurons are compared with the hypothetical motor output which presumably reaches MNs. From a computational point of view, this is a complex problem as the two activations that belong to self and the other must be maintained spatially or temporally separated in the mirror system, and compared. A computationally more plausible solution involves a forward model, by which the motor signals are converted into sensory predictions, which is then compared with the sensory input due do the observed action. This relieves the MNs of undertaking dual representation task. In his forward–inverse model architecture Demiris and Johnson [20] realized this mechanism as a computational framework to estimate the action of a demonstrator. However note that this is not a realization of Gallese and Goldman [33] suggestion since the comparison takes place in the sensory domain, and the result of this comparison leads to the mirror code that represents the observed action. The conception of Gallese and Goldman [33] employs MNs as motor decoders and necessitates a higher level decision system to do the intention decoding. Therefore this proposal is different from the intention decoding or understanding role of mirror neurons via direct matching (L3 of Section 2). In particular, it does not claim an automatic understanding by the mere firing of MNs but require an additional decision and search system. Mental State Inference (MSI) model realizes a similar search mechanism in computer simulation lending support for the computational plausibility of such search functionality [69]. However there is a fundamental difference how MNs are employed. In Gallese and Goldman's proposal MNs are motor decoders and do not have direct motor control role. On the contrary, the MSI model assigns the control role of sensory forward prediction for MNs to support fast motor execution (see Fig. 3). Similarly in the model of Demiris and Johnson [20] a control role, i.e. the inverse modeling role is envisioned for the MNs.

Raos et al. [72] indicate that action understanding extends beyond MNs, and a broad mechanism, i.e. "mental simulation" of action rather than "mirroring" is central to action understanding (see Fig. 4). This is in contrast with the proposal of Gallese and Goldman, as their MNs are central to matching a mentally simulated act. How might the proposal of Raos et al. [72] work in computational terms? One possibility is to have mental simulation generate the MN activity. This will in no way conflict with the published data on mirror neurons, because the main tool used in those studies for analysis is correlation (not causation).

If we have a simulation mechanism it can be used to obtain a motor code compatible with the observed action. Here is how. Pick a candidate motor code (Y), mentally simulate its execution and the sensory feedback that would be perceived. This would yield a series of virtual sensory signals (S_{img}). Now this signal can be compared with the actual signals (S_{actor}) thereby yielding an error for the motor code, Y. Thus, among a set of candidate motor codes the one that minimizes the error between the simulated sensory input and the actual sensory input can be selected through an appropriate mechanism. If the motor space is continuous, a gradient descent type of method may be used to reach to the motor code that minimizes the error. In this scenario, the representation of the candidate motor code or the result of recognition could be associated with MN activity (only the latter is depicted in Fig. 4). This model predicts that the temporal activity of MNs corresponds to the search through the motor code space, converging on to an answer as more and

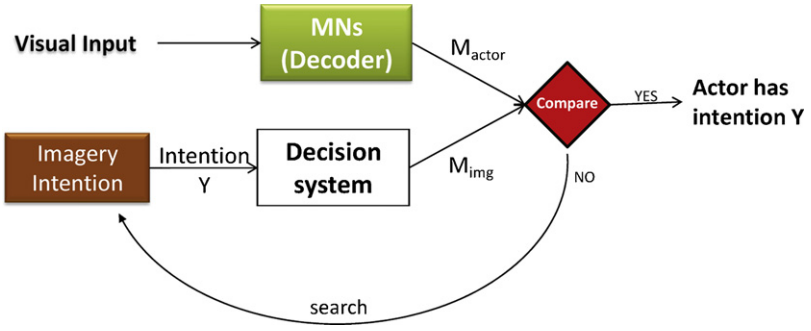

Fig. 2. Depicting the use of mirror neurons in the conceptual model of Goldman and Gallese [33] for intention understanding.

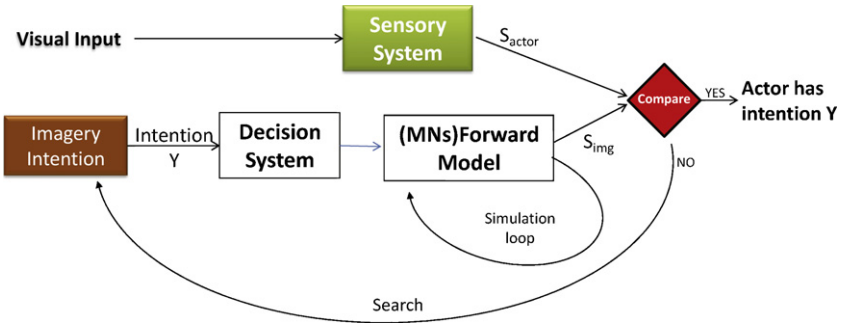

Fig. 3. Depicting the intention understanding and sensory forward prediction role of MNs in the computational model of Oztop et al. [69].

more of the action became available (e.g. as grasping action come close to completion). This model would also somewhat fit with Csibra's [17] view of the mirror neuron system, where MNs report the results of an understanding process (here, though the process is not understanding, but decoding of the motor code). Computationally, the model of Demiris and Johnson [20] can be seen as the realization of this mechanism in a parallel search architecture, where the simulation loop is replaced by a set of inverse models.

5.1. Do we read too much into correlation?

The above discussion underlines an important danger about building theories based only upon correlation analysis. The current data cannot distinguish between three possible architectures involving mirror neurons: conceptual models of Gallese and Goldman [33] and Raos et al. [72] and the computational model of Oztop et al. [69]. TMS and lesion studies may have some power in overcoming the difficulties associated with correlation analysis. However, lesion studies can at most provide necessary conditions for some brain regions (region A is necessary for some function B), and usually suffer from spatial resolution limitations, especially in the case of human patient studies. In order to prove sufficient conditions (spatiotemporal neural activities in region A is sufficient to cause function B), we need to develop an entirely different causal technique [48]. One possible example of that was recently developed as a Decoded Neuro-Feedback (DecNef) method [82]. If one can cause "understanding of actions" by inducing a specific spatial neural activity in the mirror neuron system by fMRI DecNef, we will be in a position to prove the causal role of MNs.

6. Distinguishing computational and conceptual modeling

In this section we emphasize the difference of conceptual and computational models. At one extreme the statement "an observed action is represented in motor terms" may define a conceptual model but in developing a computational model, a set of strict and unambiguous specifications must be followed. In particular, the notions of *what is matched* and *how it is represented* as in Section 2 must be precisely defined. Furthermore the observation must be also explained in the model specification: is it the input a set of time varying retinal activations? Or we assume some part of the brain recognizes the hand, arm, etc. and transforms the sensory input into a standard form so that a compact representation of a moving hand is available to the MNs? What about the temporal aspects of the observed act? For example, we noted above that the association between self generated motor code and the observed responses may yield MN-like units. This is true, and computationally plausible if the observed response is taken as static snapshots. However, if one tries to incorporate the temporal aspect of the action then the task becomes formidably difficult [93]. In the literature, we have several conceptual models at different levels of complexity that talk about MNs at the event-level [e.g. 52,53,58], but, unfortunately, there are no recent computational models at the trajectory level except those that extend the model of [65] in several ways [8,9]. Unfortunately most models of the role of MNs in *human social cognition* are purely conceptual, focusing on *where* they are located rather than on *how* they function. In the next section, we will present the Mental State Inference (MSI) Model of Oztop et al. [69] in relation to the conceptual models of Gallese and Goldman [33] and Raos et al. [72], which both have certain parallels with the MSI model.

7. Internal models and mirror neurons

Although we have indicated that direct matching (corresponding to inverse modeling at some control level) is the prime focus of the neuroscientific community for mirror neurons, there are a few proposals that involve MNs in forward modeling [13,44]. The idea being the mirror code is used to generate a visual prediction in the Superior Temporal Sulcus (STS) where neurons have been found with selectivity for biological movement (e.g. of arms, whole body) for comparison. Miall [58] suggested extending the aforementioned conceptual model by including the cerebellum. He proposed that the forward and inverse computations required can be carried out by the cerebellum and the posterior parietal cortex. These proposals require F5-STS act as forward-and-inverse models that can be thought to be analogous to the workings of the MOSAIC model [37]. However, these conceptual models grossly oversimplify the computational problem, failing to specify how the retinal image is transformed to a manageable form and size in STS and how the temporal aspects of motor control factor in the motor representation. Moreover, for inverse and forward models related to the whole body, learning is not straightforward for one cannot completely observe all of one's own body via direct vision – suggesting the need for models that take fuller account of proprioception. Kilner et al. [53] formalize the proposal of Carr et al. [13] by suggesting the application of Empirical Bayesian inference for inverting the generative model that captures the (*motor command* → *generated stimuli*) mapping in the F5-STS complex. This proposal is conceptually sound, but hard to evaluate as no implementation was provided. The MSI model preceding this work, in fact, realized most of these ideas in a biologically plausible yet computationally feasible way.

7.1. Mental State Inference model

The Mental State Inference (MSI) model illustrated in Fig. 5 was developed to answer the question of how an agent can obtain an estimate of the demonstrator's goal or intention based on visual observation [69]. This model is similar to the proposal of Gallese and Goldman [33] in that the goal of the demonstrator is extracted, but no *mirroring* mechanism is assumed. Instead, it

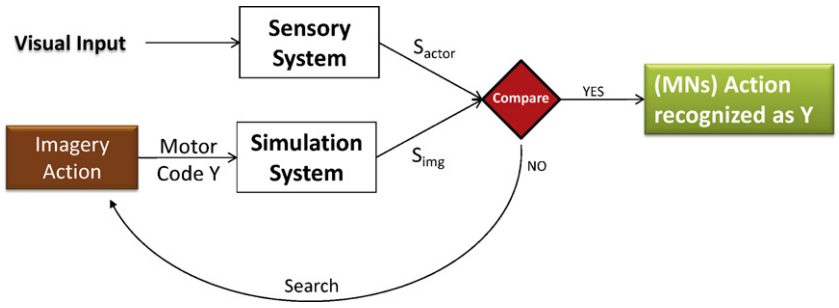

Fig. 4. Depicting the use of a simulation system for decoding the observed actions according to the conceptual model of Raos et al. [72].

takes the *mental simulation* as the core mechanism for intention understanding, making it similar to the proposal of Raos et al. [72]. According to the MSI model, the dual activation of mirror neurons is explained by these two processes: (1) automatic engagement of Mental State Inference during action observation, and (2) the forward prediction task undertaken by the mirror neurons for motor control during action execution.

In the MNS series of models for the mirror neuron system [8,9,65], self-observation was key for the development of mirror neurons. Although these models hold that mirror neurons had to be involved in motor control, it was not computationally shown how this function would be carried out. On the other hand in the MSI model a sensory forward model role was considered to be the computational analog of mirror neurons [69]. The MSI model attempts to give an account of how Mental State Inference can be realized with one's own motor system once a forward prediction capability is available for motor control.

The model first proposes a visual feedback control circuit and shows how parts of this circuit can be utilized for inferring others' intentions. The postulated movement control proceeds as the following (see Fig. 5). The parietal cortex extracts visual features relevant for the control of a particular goal-directed action (x, the control variable) and relays this information to the premotor cortex (for easier comprehension, one may assume that x is the distance,

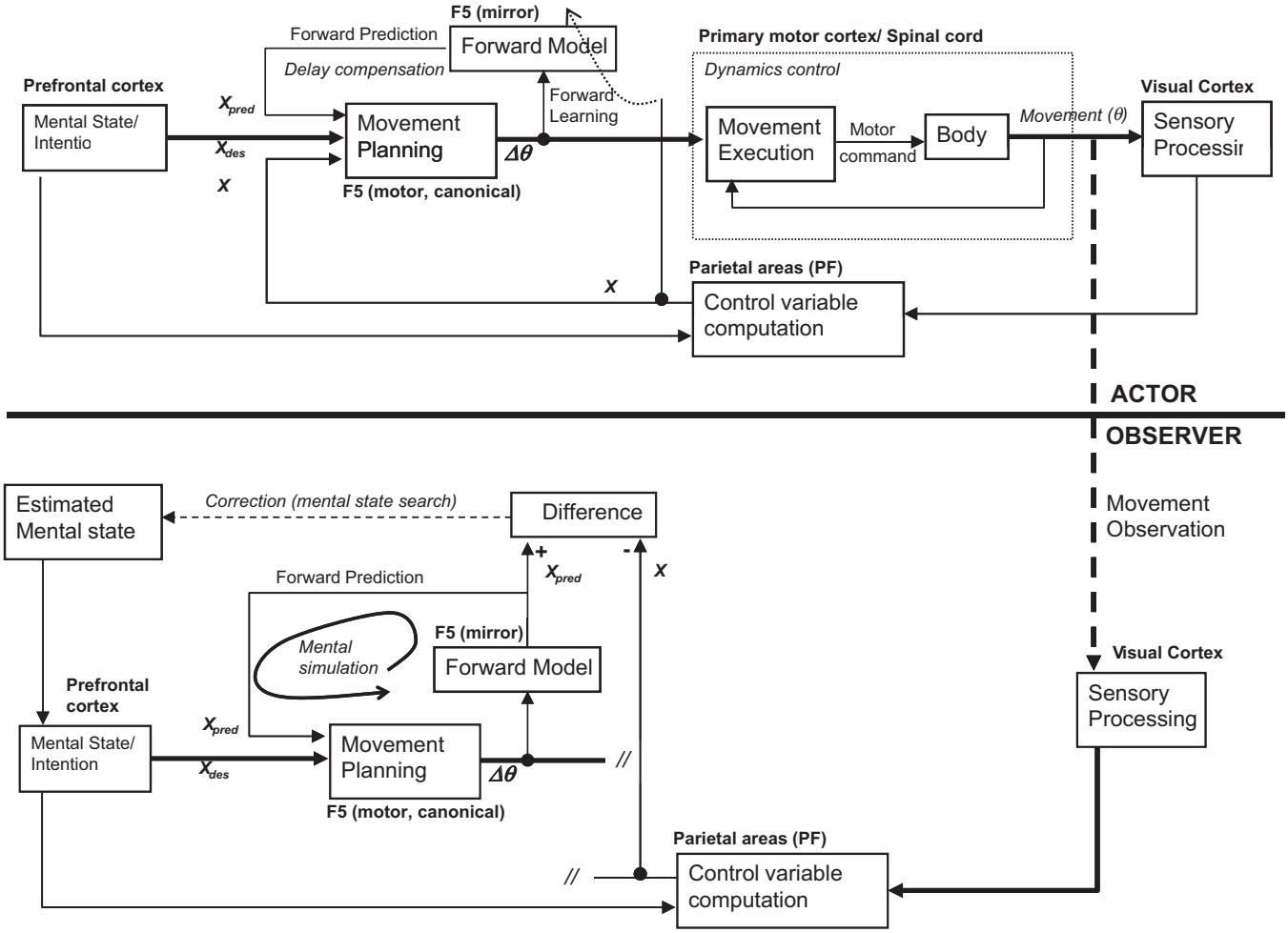

Fig. 5. The functioning of the Mental State Inference (MSI) model is illustrated.

Source: Adopted from [69].

for the task of reaching to a desired position). The premotor cortex computes a motor code (m) to bring the parietal cortex output (x) to the desired neural code (x_{des}) relayed by the prefrontal cortex (for fulfilling an intention). The desired change generated by the premotor cortex is used by dynamics related motor centers (cerebellum, brain stem and spinal cord) for execution. During observation, the forward model (F5 mirror neurons) continuously creates predictions as to what will be the next sensory input based on input motor code. Observing agent starts with a guess of the mental state of the demonstrator, and simulates the action that is appropriate for that mental state using the forward model in a mental simulation loop. This generates a sequence of signals as mirror neuron output ($M_i = m_{i1}, m_{i2} \ldots$). In parallel, the perceptual system of the observer is engaged in processing the demonstrator's action, thereby generating a sequence of signals ($A = a_1, a_2, a_3 \ldots$) as the action ensues. Mental simulation loop is very fast (because no actual movement is involved); so, many mental states ($M_1, M_2, \ldots$) can be tried generating a set of predicted sensory signal sequences while the observer is moving. Therefore a search mechanism can be implemented to find the mental state k that gives the best match, i.e. $M_k \sim A$. In [69] two search mechanisms were implemented: exhaustive search for discrete mental states, and gradient descent for continuous mental states. The key issue that made the model generalize from the knowledge applicable to self-object interactions to other-object interactions was the definition of the control in an object centered way. That is the control signals computed are always with respect to a target object. In grasping an object (e.g. a hammer), the time course of the angle between palm normal and the principle axis of the target object and distance to it suffices to understand the intention of an agent in grasping the object [69]. In the work of Umilta et al. [94] mirror neuron firings were shown to positively correlate with the reverse pliers motion rather than the finger movements. It was inferred that these neurons encode the intention of the act since they fire when gripping of the object occurs regardless of the effector (finger or the pliers). However, this could just well be explained by the fact these neurons predict the sensory change (i.e. the decreasing distance between the object and effector) during the execution and observation of the gripping of the object regardless of the agent and the effector used. In this sense we see the work of Umilta et al. [94] as a verification of one of the predictions of the MSI model (see also [4] in relation to the distalization of the end effector).

The recent data on cortico-spinal mirror neurons indicate that mirroring in monkeys is not limited to object-hand interactions; but, index and thumb finger flexion miming a precision pinch can also trigger mirror activity [55]. Furthermore, where this activity is manifested as an increase in the firing rate of some cortico-spinal MNs, in others a decrease was observed. When viewed from a MSI model point of view, the increase/decrease modulation is natural as not all control parameters need to follow a monotonic trajectory. The lack of the necessity of a goal object for mirror activity in monkeys is a significant finding, and must be investigated further. This phenomenon can be naturally accommodated by the MSI model by limiting the control parameters to the hand, e.g. distance between the finger tips.

8. Development of mirror neurons

Despite the growing number of reports on adult mirror neuron systems, data on development of the mirror system for both human and other primates is scarce. It is known that the macaque mirror neuron system is adaptive: responses to novel actions can be acquired through repeated experience. For example, some mirror neurons respond to both execution and viewing of ripping a sheet of paper, which is not in the ecological repertoire of wild monkeys [54]. Furthermore, as we have seen, some mirror neurons that were originally responsive to precision pinch observation became responsive to use of pliers for picking up objects [94]. These macaque data suggest mirror neuron responses are learned rather than innate. In addition, Catmur et al. [15] have shown that (the manifestation of) the action-observation matching system in humans can be altered (see also Catmur, this issue). Humans show event-related muscle-specific responses to transcranial magnetic stimulation (TMS) over the motor cortex during the observation of actions. Observation of an action that involves a particular muscle elicits a stronger response in the very same muscle of the observer compared to an incongruent action observation [see 24]. This response is generally associated with the human mirror system. The experiment of Catmur et al. [15] asked subjects to perform finger actions (little and index finger) in response to presented finger movements (little and index finger). One group was asked to perform congruent movements, while the other group was asked to perform the incongruent movement upon the observation of the presented action. After training, both groups were tested for their response to observed finger movements. The congruent group retained the mirror property: observation of the index [little] finger movement, compared with the little [index] finger observation, generated a higher response in the muscle that actuates the index [little] finger. This was completely reversed in the incongruent group; the sensorimotor congruency during training adapted the mirror system so that coupling between the observed and executed movement was altered. Therefore, what makes observed actions to be mirrored in motor terms is due to the obvious fact that motor execution always creates congruent visual stimuli (a motor command to move the index finger moves the index finger) which is experienced over the lifetime of the organism.

These data confirm the plastic nature of the mirror system both in humans and monkeys. However they do not rule out a rudimentary mirror system at birth. This view is adopted often by the proponents of the idea that a direct link between mirror neurons and imitation exists, with the logic that infant imitation must be due to an innate mirror neuron system. There have been several experimental studies that aimed at detecting the infant mirror system, which involved subjects of 14 months old [57], 9 months old [84] and 6 months old infants [64]. Although the results from these studies are important for understanding the development of MNs, this age range is not really sufficient to answer the innateness question; because (grasp related) MNs may develop in the critical phase of 2–6 months where reaching is transformed into voluntary grasping [65–67]. In these studies high resolution EEG was used to detect mirror activity (see also Vanderwert, Fox, Ferrari, this issue). Classically, the EEG marker for mirror neuron activity in adult humans is taken as the suppression in the mu frequency band (8–13 Hz) over the sensorimotor and motor areas. At rest, the mu frequency band displays high amplitude oscillations due to the synchronized firing of neurons; but, when the subjects perform an action, or observe movements, the power of the mu band attenuates due to desynchronization [57]. The rest frequency and the amplitude of the mu rhythm in infants is lower compared to the adults' [see 64 and citations therein]. All of the aforementioned studies reported a suppression in the mu frequency band both in observation and execution conditions suggesting that the mirror neuron system may be functioning in 6 months and older infants. Of these only Nystrom [64] compared the mu band activity of adults to those of infants. The adult group showed a significant difference between the goal-directed and the non-goal directed action observation conditions. Although, the infants showed the same pattern of desynchronization as adults between conditions, the difference did not reach significance which may indicate a non-mature mirror system at this age. However, when ERPs 0.5 s before contact, or at contact were used the infant difference between conditions

also become significant. A final indication of the possible immaturity of the infant mirror system is the significant activation found in infants for observing a moving-dot compared to the baseline of observing a static dot, which was not the case for adults.

Taken together, the data indicate that mirror system is adaptive and may start to function as early as 6 months post-birth, albeit at a rudimentary level. However, the question of innateness is still largely unanswered as the infant (grasp related) mirror system may have plenty of data to develop in the critical phase of 2–6 months where infants learn to grasp. Our view is that 'learning to grasp' facilitates 'learning to mirror', and both develop through the first year of life. The mirror systems for other skills (e.g. clapping) should similarly develop along with the development of the skill. As a final note on mirror neuron system development, it must be underlined that the computational requirements for developing a mirror system for different actions may differ vastly; for example manual actions is open to self observation whereas facial gestures are not. Thus, development of a facial mirror system requires a mechanism other than the self observation [68].

9. Evolution of the language-ready brain

The mirror system hypothesis (MSH) of the evolution of the language-ready brain has developed over the years [2,3,74] to embrace a wide body of data. This is not the place to review those developments or the details of the current version of the model. Instead, we simply summarize the stages in biological evolution of the brain that the hypothesis posits, noting that it progresses from a mirror system for manual actions posited for our last common ancestor with monkeys 25 mya by a posited series of expansions which link mirror systems able to detect more nuances of actions to an expanding range of capabilities in other neural circuits:

By 5–7 mya, the last common ancestor with chimpanzees had developed a neural capability for simple imitation of manual actions. This proved sufficient for troupes of conspecifics to develop small sets, 10 or so, of manual gestures used for communication, perhaps through a process of ontogenetic ritualization [91], possibly accompanied by social learning [5]. Subsequently, according to MSH, our ancestors (but not those of other apes) developed brain mechanisms supportive of (i) the ability to recognize another's performance as a set of familiar movements, (ii) complex action recognition (more generally) as the ability to recognize that another's performance reaches observed subgoals by combining actions which can be imitated more or less crudely by variants of actions and/or movements already in the repertoire; and (iii) mechanisms for complex imitation which exploit these recognition processes to attempt to approximate the performance on this basis, with increasing practice yielding increasing skill. Studies linking such developments to the archeological record of stone tool making suggest that the transition to brains supporting complex imitation may have begun with the transition from Oldowan to Acheulean technology (ca. 1.6 mya), but was only completed during the late Acheulean (ca. 0.7–0.25 mya) before the emergence of *Homo sapiens* ca. 0.2 mya [1,85–87].

Human fMRI studies have consistently shown that cerebro-cerebellar communication loops play essential role in representing internal models of various tools, and Broca's area and lateral parts of the cerebellum are involved in this loop [42,45,90,96]. The overlap of brain activity for perceiving language and using tools in Broca's area [41] suggests that language and tool use share computational principles for processing complex hierarchical structures common to these two abilities. Furthermore, Broca's area being a possible homologue of the monkey ventral premotor cortex where mirror neurons are located suggests that neural processes for manipulation of complex hierarchical structures may have existed in primates, and are exapted to support the change from protolanguage to language.

It is important to distinguish protolanguage (whether signed, spoken or both) from language – the suggestion being that the former can develop an open-ended lexicon whereas the latter combines an open-ended lexicon with a grammar that supports the ability to combine words to express new meanings to meet the demands of the current situation. According to MSH, the path to protolanguage rested on biological selection, whereas the emergence of language from protolanguage did not require further changes in the human genome but rather rested on tens of millennia of cultural evolution [3, Chapter 10].

The capacity for protolanguage rested on three further changes of mirror systems in relation to other portions of the brain: The first supported pantomime by exapting complex action recognition for communicative actions. The crucial point here is that this supported the change from the limited repertoire of manual communicative gestures we see now in groups of apes to an open-ended semantics. The problem with pantomime, though, is that it is costly to perform and hard to disambiguate. This created the adaptive pressure for the evolution of brain structures to support early forms of *protosign*, a manual communication system, based in part on conventionalization of pantomimes to yield an open-ended semantics. To address the fact that nonhuman primates lack the vocal control to imitate human speech sounds, MSH posits that protospeech rested on the "invasion" of the vocal apparatus by collaterals from the communication system based on the adaptive pressure to emulate and expand upon protosign semantics. The biological bases for protosign and protospeech then evolved together in an expanding spiral to yield a language-ready brain. MSH is thus a semantics first, speech second theory. Other theories posit that control of the vocal apparatus came first (much as birds and whales have "song" without words) and semantics came later [18,27,59], but in either case an account has to be given as to how humans acquired a mirror system for speech articulation that complements the ancestral mirror system for manual action.

10. Discussion

Our discussion so far indicates that the key to understanding the function of MNs and the mechanisms that facilitate that function is governed by our knowledge on the coding of MNs and how evolution changed and augmented MNs.

What do MNs encode (during action execution) and decode (during action observation)? As brain imaging cannot offer much help in this endeavor due to the gross temporal and/or spatial smearing of the neural activity; we hold that most direct information can be obtained via neurophysiology focused on *the population level temporal activation* of MNs.

10.1. Experiments

From the macaque literature we know the notion of strong versus weak congruence of the MNs. Some neurons show high congruence for observation and execution coupling whereas others broadly link actions and observations (e.g. a broadly congruent MN may fire for only precision pinch execution but may respond to observation of any type of grasping). Based on this, one straightforward suggestion is that MNs encode *features* of an observed action. A strictly congruent neuron encodes a feature unique to just one subclass of actions being tested; a broadly congruent neuron codes a feature shared by a larger class of actions. The differential activation of MNs for different actions allows a distributed representation (population level encoding) of observed actions as the vector of firing of an ensemble of MNs. In this view, the MNs are envisioned at the same hierarchical level of feature encoding (but

with different level of responses for the actions in the hierarchy level), somewhat analogous to subscribing to one of the L1, L2, L3 views of Section 2. The other possibility (*heterogeneous encoding view*) is to envision MNs to be heterogeneous with respect to L1, L2, L3 and other properties, so that for example, some MNs may code motor parameters, others may code goals, and some can even code both (see also Casile, et al. [14] for more on this). The support for this view comes from the parietal mirror neurons of Fogassi et al. [28] which may code the conjunction of proximal actions and distal goals. This view can explain the multiplicity of functions attributed to mirror neurons in the literature [73]. It is important to remember that monkeys do not imitate, so if the latter view is correct, the population balance between L1 and L3 and other properties must show shifts across species and with experience. The proposal of Casile et al. [14] that MNs can represent a system integrating behaviorally relevant characteristics of actions which may be experience and subject dependent lends support to the heterogeneous encoding view of MNs outlined above.

Although, at the beginning of this section we indicated that brain imaging is limited in helping uncover mirror neuron function, novel advanced methods based on encoding-decoding metaphor and the neural inception technique of Shibata et al. [82] may somewhat ameliorate these limitations. The relationships between decoding and encoding have been a important topic in fMRI multi-voxel decoding literature. Kamitani and co-workers demonstrated that very fine information such as orientation of black and white images can be decoded from fMRI BOLD signals through machine learning techniques [47,60]. In these (conventional) fMRI decoding techniques, maps from BOLD signals to some useful information that should be extracted from the brain were directly obtained based on (brain signal, stimuli) data pairs. Recently, a new technique of "decoding by encoding" was proposed and successfully applied to find best matching movies from BOLD signals [62,63]. In this study, first a good quantitative model of encoding BOLD signal for each voxel from the visual stimuli was developed based on earlier neurophysiological studies. This encoding model was later utilized in decoding by using a Bayesian approach. This paradigm can be applied to MN research by building the encoding model based on brain activity during execution, and later decoding the different actions observed based on this model. By controlling the overlapping features of the executed and observed actions, the true encoding used by MNs may be inferred.

Using the reasoning that observation of skill performance improves the execution of the skill, one can also utilize the inception technique of Shibata et al. [82] for investigating the MN system using fMRI imaging. The subjects can be incepted with mirror responses corresponding to the observation of a novel action (X). Subjects, before and after the inception, can then be asked to execute a set of tasks (Y1, Y2, ...Y*n*) that have varying overlaps with the action that was used in incepting the MN activity. The actions that are performed better after the inception can be used to infer the features that were encoded by the MNs (provided that Ys were chosen so that their features have varying overlap with those of X).

10.2. Models

The wide interest that MNs received as bases for cognitive abilities such as imitation, action understanding, intention attribution and mental state often hides the very basic questions that need to be answered in order to build a solid theory of MNs. One might call the current state of MN research as a 'shadow theory' for which the research effort seems to focus on the properties of it, rather than finding the source of the shadow. Neurophysiology is indispensible for this endeavor when used in conjunction with computational modeling that verifies a given conceptual theory and produce testable predictions. To be productive in building the theory, MN research needs to leave the correlation based thinking and move towards the causal relations between events and brain activations. The current computationally oriented proposals for mirror neuron function (in monkey) are related to the self action of the monkey. In the model of Demiris and Johnson [20] or Haruno et al. [37] the inverse model outputs can be associated with MN activity. In the model of Oztop et al. [69], the sensory forward prediction output is associated with MN activity. The latter proposals imply a control role for the mirror neurons. Another hypothesis is that MNs are used for action monitoring, and not necessarily involved in control per se. In this view, MNs are irrelevant to the successful execution of an over-learned action under normal circumstances; but, may indicate a failure of execution, or near-completion of a goal by an unintended action (e.g. due to a perturbation) thereby facilitating new action formation to achieve the goal [8]. The beautiful part of all three computational suggestions is that they can be *tested* experimentally. For example, for the latter model, an experiment that checks how MN activity changes when random perturbations are applied during self-actions can be designed. In fact, there are very few studies that investigate the MNs in relation to execution. Of those, the study of [29] investigated the effect of the inactivation of MNs via muscimol injection. The results showed that the grasp related mirror neurons are not so much needed for the planning and execution of the grasping action (a mere slowing down was observed). This data is compatible with both models of ACQ ("what did I just do") [8] and MSI (Mental State Inference) models [69].

10.3. Evolution

With the discussion above, we by no means claim that the role of MNs in cognitive functions can be studied only with the help of computational modeling. We know that monkeys have mirror neurons and for the most part they do not demonstrate the cognitive functions that are attributed to MNs. Therefore the missing link has to be explained with the evolutionary changes that transformed or augmented the monkey mirror system so to allow it to undertake higher cognitive functions. In fact, our review of the mirror system hypothesis for the evolution of the language-ready brain has offered for further study one possible path whereby the human brain's capacity for imitation and language required not only evolution of the human mirror system from/on top of the mirror system of our macaque-human last common ancestor but also massive changes in the network of regions with which the system interacts. Many challenges remain to test and refine this theory while expanding it to provide satisfactory approaches to other social functions such as theory of mind and emotional interaction.

Acknowledgement

This material is based in part on work supported by the National Science Foundation under Grant No. 0924674 (Michael A. Arbib, Principal Investigator). We thank Prof. Akira Murata for his useful comments and pointers on the anatomical connections of area F5. Author MK was supported by the Strategic Research Program for Brain Sciences of Japanese MEXT.

References

[1] M.A. Arbib, From mirror neurons to complex imitation in the evolution of language and tool use, Annual Review of Anthropology 40 (2011) 257–273.
[2] M.A. Arbib, From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics, The Behavioral and Brain Sciences 28 (2005) 105–124 (discussion 125-167-105-124; discussion 125-167).
[3] M.A. Arbib, How the Brain Got Language: The Mirror System Hypothesis, Oxford University Press, New York, 2012.
[4] M.A. Arbib, J. Bonaiuto, S. Jacobs, S.H. Frey, Tool use and the distalization of the end-effector, Psychological Research 73 (2009) 441–462.

[5] M.A. Arbib, K. Liebal, S. Pika, Primate vocalization, gesture, and the evolution of human language, Current Anthropology 49 (2008) 1053–1076.
[6] L. Aziz-Zadeh, T. Sheng, S.L. Liew, H. Damasio, Understanding otherness: the neural bases of action comprehension and pain empathy in a congenital amputee, Cerebral Cortex 22 (2012) 811–819.
[7] A. Billard, M.J. Mataric, Learning human arm movements by imitation: evaluation of a biologically inspired connectionist architecture, Robotics and Autonomous Systems 37 (2001) 145–160.
[8] J. Bonaiuto, M.A. Arbib, Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons, Biological Cybernetics 102 (2010) 341–359.
[9] J. Bonaiuto, E. Rosta, M. Arbib, Extending the mirror neuron system model, I. Audible actions and invisible grasps, Biological Cybernetics 96 (2007) 9–38.
[10] G. Buccino, F. Binkofski, G.R. Fink, L. Fadiga, L. Fogassi, V. Gallese, R.J. Seitz, K. Zilles, G. Rizzolatti, H.J. Freund, Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study, European Journal of Neuroscience 13 (2001) 400–404.
[11] R.W. Byrne, Imitation as behaviour parsing, Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 358 (2003) 529–536.
[12] D.P. Carey, D.I. Perrett, M.W. Oram, Recognizing, understanding, and producing action, in: M. Jeannerod, J. Grafman (Eds.), Handbook of Neuropsychology, Action and Cognition, vol. 11, Elsevier, Amsterdam, 1997, pp. 111–130.
[13] L. Carr, M. Iacoboni, M.-C. Dubeau, J.C. Mazziotta, G.L. Lenzi, Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas, PNAS 100 (2003) 5497–5502.
[14] A. Casile, V. Caggiano, P.F. Ferrari, The mirror neuron system: a fresh view, Neuroscientist 17 (2011) 524–538.
[15] C. Catmur, V. Walsh, C. Heyes, Sensorimotor learning configures the human mirror system, Current Biology 17 (2007) 1527–1531.
[16] T. Chaminade, E. Oztop, G. Cheng, M. Kawato, From self-observation to imitation: visuomotor association on a robotic hand, Brain Research Bulletin 75 (2008) 775–784.
[17] G. Csibra, Mirror neurons and action observation. Is simulation involved? What do mirror neurons mean? Interdisciplines 8 (2005) http://www.interdisciplines.org
[18] C. Darwin, The Descent of Man and Selection in Relation to Sex, John Murray, London, 1871.
[19] F. de Vignemont, P. Haggard, Action observation and execution: what is shared? Social Neuroscience 3 (2008) 421–433.
[20] Y. Demiris, M. Johnson, Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning, Connection Science 15 (2003) 231–243.
[21] G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, G. Rizzolatti, Understanding motor events—a neurophysiological study, Experimental Brain Research 91 (1992) 176–180.
[22] I. Dinstein, Human cortex: reflections of mirror neurons, Current Biology 18 (2008) R956–R959.
[23] I. Dinstein, C. Thomas, M. Behrmann, D.J. Heeger, A mirror up to nature, Current Biology 18 (2008) R13–R18.
[24] L. Fadiga, L. Craighero, Electrophysiology of action representation, Journal of Clinical Neurophysiology 21 (2004) 157–169.
[25] L. Fadiga, L. Craighero, G. Buccino, G. Rizzolatti, Speech listening specifically modulates the excitability of tongue muscles: a TMS study, European Journal Neuroscience 15 (2002) 399–402.
[26] A.H. Fagg, M.A. Arbib, Modeling parietal-premotor interactions in primate control of grasping, Neural Networks 11 (1998) 1277–1303.
[27] W.T. Fitch, The Evolution of Language, Cambridge University Press, Cambridge, 2010.
[28] L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, G. Rizzolatti, Parietal lobe: from action organization to intention understanding, Science 308 (2005) 662–667.
[29] L. Fogassi, V. Gallese, G. Buccino, L. Craighero, L. Fadiga, G. Rizzolatti, Cortical mechanism for the visual guidance of hand grasping movements in the monkey—a reversible inactivation study, Brain 124 (2001) 571–586.
[30] L. Fogassi, V. Gallese, G. Dipellegrino, L. Fadiga, M. Gentilucci, G. Luppino, M. Matelli, A. Pedotti, G. Rizzolatti, Space coding by premotor cortex, Experimental Brain Research 89 (1992) 686–690.
[31] L. Fogassi, V. Gallese, L. Fadiga, G. Rizzolatti, Neurons responding to the sight of goal-directed hand/arm actions in the parietal area PF (7b) of the macaque monkey, in: 28th Annual Meeting of Society for Neuroscience, Los Angeles, 1998.
[32] V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, Action recognition in the premotor cortex, Brain 119 (1996) 593–609.
[33] V. Gallese, A. Goldman, Mirror neurons and the simulation theory of mind-reading, Trends in Cognitive Sciences 2 (1998) 493–501.
[34] V. Gallese, C. Keysers, G. Rizzolatti, A unifying view of the basis of social cognition, Trends in Cognitive Sciences 8 (2004) 396–403.
[35] A.P. Georgopoulos, Neuronal population coding of movement direction, Science 233 (1986) 1416–1419.
[36] M. Gerbella, A. Belmalih, E. Borra, S. Rozzi, G. Luppino, Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5, Brain Structure and Function 216 (2011) 43–65.
[37] M. Haruno, D.M. Wolpert, M. Kawato, MOSAIC model for sensorimotor learning and control, Neural Computation 13 (2001) 2201–2220.
[38] C. Heyes, Where do mirror neurons come from? Neuroscience and Biobehavioral Reviews 34 (2010) 575–583.
[39] G. Hickok, D. Poeppel, Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language, Cognition 92 (2004) 67–99.
[40] G. Hickok, D. Poeppel, Opinion—the cortical organization of speech processing, Nature Reviews Neuroscience 8 (2007) 393–402.
[41] S. Higuchi, T. Chaminade, H. Imamizu, M. Kawato, Shared neural correlates for language and tool use in Broca's area, Neuroreport 20 (2009) 1376–1381.
[42] S. Higuchi, H. Imamizu, M. Kawato, Cerebellar activity evoked by common tool-use execution and imagery tasks: an fMRI study, Cortex 43 (2007) 350–358.
[43] M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J.C. Mazziotta, G. Rizzolatti, Grasping the intentions of others with one's own mirror neuron system, PLoS Biology 3 (2005) e79.
[44] M. Iacoboni, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, G. Rizzolatti, Cortical mechanisms of human imitation, Science 286 (1999) 2526–2528.
[45] H. Imamizu, T. Kuroda, S. Miyauchi, T. Yoshioka, M. Kawato, Modular organization of internal models of tools in the human cerebellum, Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 5461–5466.
[46] S. Jacobs, C. Danielmeier, S.H. Frey, Human anterior intraparietal and ventral premotor cortices support representations of grasping with the hand or a novel tool, Journal of Cognitive Neuroscience 22 (2010) 2594–2608.
[47] Y. Kamitani, F. Tong, Decoding the visual and subjective contents of the human brain, Nature Neuroscience 8 (2005) 679–685.
[48] M. Kawato, From 'understanding the brain by creating the brain' towards manipulative neuroscience, Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 363 (2008) 2201–2214.
[49] M. Kawato, Internal models for motor control and trajectory planning, Current Opinion in Neurobiology 9 (1999) 718–727.
[50] M. Kawato, D. Wolpert, Internal models for motor control, SensoryGuidance of Movement 218 (1998) 291–307.
[51] M. Kawato, K. Furukawa, R. Suzuki, A hierarchical neural network model for control and learning of voluntary movement, Biological Cybernetics 57 (1987) 169–185.
[52] C. Keysers, D.I. Perrett, Demystifying social cognition: a Hebbian perspective, Trends in Cognitive Sciences 8 (2004) 501–507.
[53] J.M. Kilner, K.J. Friston, C.D. Frith, Predictive coding: an account of the mirror neuron system, Cognitive Processing 8 (2007) 159–166.
[54] E. Kohler, C. Keysers, M.A. Umilta, L. Fogassi, V. Gallese, G. Rizzolatti, Hearing sounds, understanding actions: action representation in mirror neurons, Science 297 (2002) 846–848.
[55] A. Kraskov, N. Dancause, M.M. Quallo, S. Shepherd, R.N. Lemon, Corticospinal neurons in macaque ventral premotor cortex with mirror properties: a potential mechanism for action suppression? Neuron 64 (2009) 922–930.
[56] Y. Kuniyoshi, Y. Yorozu, M. Inaba, H. Inoue, From visuo-motor self learning to early imitation—a neural architecture for humanoid learning, in: International Conference on Robotics & Automation, IEEE, Taipei, Taiwan, 2003.
[57] P.J. Marshall, T. Young, A.N. Meltzoff, Neural correlates of action observation and execution in 14-month-old infants: an event-related EEG desynchronization study, Developmental Science 14 (2011) 474–480.
[58] R.C. Miall, Connecting mirror neurons and forward models, Neuroreport 14 (2003) 2135–2137.
[59] S. Mithen, The Singing Neanderthals: The Origins of Music, Language, Mind & Body, Weidenfeld & Nicholson, London, 2005.
[60] Y. Miyawaki, H. Uchida, O. Yamashita, M.A. Sato, Y. Morito, H.C. Tanabe, N. Sadato, Y. Kamitani, Visual image reconstruction from human brain activity using a combination of multiscale local image decoders, Neuron 60 (2008) 915–929.
[61] A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, G. Rizzolatti, Object representation in the ventral premotor cortex (area F5) of the monkey, Journal of Neurophysiology 78 (1997) 2226–2230.
[62] T. Naselaris, K.N. Kay, S. Nishimoto, J.L. Gallant, Encoding and decoding in fMRI, Neuroimage 56 (2011) 400–410.
[63] S. Nishimoto, A.T. Vu, T. Naselaris, Y. Benjamini, B. Yu, J.L. Gallant, Reconstructing visual experiences from brain activity evoked by natural movies, Current Biology 21 (2011) 1641–1646.
[64] P. Nystrom, The infant mirror neuron system studied with high density EEG, Social Neuroscience 3 (2008) 334–347.
[65] E. Oztop, M.A. Arbib, Schema design, implementation of the grasp-related mirror neuron system, Biological Cybernetics 87 (2002) 116–140.
[66] E. Oztop, N.S. Bradley, M. Arbib, The development of grasping and the mirror system, in: M. Arbib (Ed.), Action to Language via the Mirror Neuron System, Cambridge University Press, New York, 2006.
[67] E. Oztop, N.S. Bradley, M.A. Arbib, Infant grasp learning: a computational model, Experimental Brain Research 158 (2004) 480–503.
[68] E. Oztop, M. Kawato, M. Arbib, Mirror neurons and imitation: a computationally guided review, Neural Networks 19 (2006) 254–271.
[69] E. Oztop, D. Wolpert, M. Kawato, Mental state inference using visual control parameters, Brain Research. Cognitive Brain Research 22 (2005) 129–151.
[70] D.I. Perrett, M.H. Harries, R. Bevan, S. Thomas, P.J. Benson, A.J. Mistlin, A.J. Chitty, J.K. Hietanen, J.E. Ortega, Frameworks of analysis for the neural representation of animate objects and actions, Journal of Experimental Biology 146 (1989) 87–113.
[71] D.I. Perrett, P.A. Smith, A.J. Mistlin, A.J. Chitty, A.S. Head, D.D. Potter, R. Broennimann, A.D. Milner, M.A. Jeeves, Visual analysis of body movements by

neurones in the temporal cortex of the macaque monkey: a preliminary report, Behavioural Brain Research 16 (1985) 153–170.
[72] V. Raos, M.N. Evangeliou, H.E. Savaki, Mental simulation of action in the service of action perception, The Journal of Neuroscience 27 (2007) 12675–12683.
[73] G. Rizzolatti, The mirror neuron system and its function in humans, Anatomy and Embryology 210 (2005) 1–3.
[74] G. Rizzolatti, M.A. Arbib, Language within our grasp, Trends in Neurosciences 21 (1998) 188–194.
[75] G. Rizzolatti, L. Craighero, The mirror-neuron system, Annual Review of Neuroscience 27 (2004) 169–192.
[76] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi, Premotor cortex and the recognition of motor actions, Cognitive Brain Research 3 (1996) 131–141.
[77] G. Rizzolatti, L. Fogassi, V. Gallese, Motor and cognitive functions of the ventral premotor cortex, Current Opinion in Neurobiology 12 (2002) 149–154.
[78] G. Rizzolatti, L. Fogassi, V. Gallese, Neurophysiological mechanisms underlying the understanding and imitation of action, Nature Reviews Neuroscience 2 (2001) 661–670.
[79] J.M.A. Roberts, Echolalia, comprehension in autistic-children, Journal of Autism and Developmental Disorders 19 (1989) 271–281.
[80] H. Sakata, M. Taira, M. Kusunoki, A. Murata, Y. Tanaka, The TINS lecture—the parietal association cortex in depth perception and visual control of hand action, Trends in Neurosciences 20 (1997) 350–357.
[81] H. Sakata, M. Taira, A. Murata, S. Mine, Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey, Cerebral Cortex 5 (1995) 429–438.
[82] K. Shibata, T. Watanabe, Y. Sasaki, M. Kawato, Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation, Science 334 (2011) 1413–1415.
[83] J.I. Skipper, H.C. Nusbaum, S.L. Small, Listening to talking faces: motor cortical activation during speech perception, Neuroimage 25 (2005) 76–89.
[84] V. Southgate, M.H. Johnson, T. Osborne, G. Csibra, Predictive motor activation during action observation in human infants, Biology Letters 5 (2009) 769–772.
[85] D. Stout, The social and cultural context of stone-knapping skill acquisition, in: V. Roux, B. Bril (Eds.), Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behaviour, McDonald Institute for Archaeological Research, Cambridge, 2005, pp. 331–340.
[86] D. Stout, Stone toolmaking and the evolution of human culture and cognition, Philosophical Transactions of the Royal Society B: Biological Sciences 366 (2011) 1050–1059.
[87] D. Stout, T. Chaminade, Stone tools, language and the brain in human evolution, Philosophical Transactions of the Royal Society B: Biological Sciences 367 (2012) 75–87.
[88] M. Taira, S. Mine, A.P. Georgopoulos, A. Murata, H. Sakata, Parietal cortex neurons of the monkey related to the visual guidance of hand movement, Experimental Brain Research 83 (1990) 29–36.
[89] M. Takada, A. Nambu, N. Hatanaka, Y. Tachibana, S. Miyachi, M. Taira, M. Inase, Organization of prefrontal outflow toward frontal motor-related areas in macaque monkeys, The European Journal of Neuroscience 19 (2004) 3328–3342.
[90] T. Tamada, S. Miyauchi, H. Imamizu, T. Yoshioka, M. Kawato, Cerebro-cerebellar functional connectivity revealed by the laterality index in tool-use learning, Neuroreport 10 (1999) 325–331.
[91] M. Tomasello, J. Call, Primate Cognition, Oxford University Press, New York, 1997, 528 pp.
[92] S. Uithol, I. van Rooij, H. Bekkering, P. Haselager, Understanding motor resonance, Social Neuroscience 6 (2011) 388–397.
[93] S. Uithol, I. van Rooij, H. Bekkering, P. Haselager, What do mirror neurons mirror? Philosophical Psychology 24 (2011) 607–623.
[94] M.A. Umilta, L. Escola, I. Intskirveli, F. Grammont, M. Rochat, F. Caruana, A. Jezzini, V. Gallese, G. Rizzolatti, When pliers become fingers in the monkey motor system, Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 2209–2213.
[95] M.A. Umilta, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, G. Rizzolatti, I know what you are doing: a neurophysiological study, Neuron 31 (2001) 155–165.
[96] K.R. Van Dijk, T. Hedden, A. Venkataraman, K.C. Evans, S.W. Lazar, R.L. Buckner, Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization, Journal of Neurophysiology 103 (2010) 297–321.
[97] D.M. Wolpert, Z. Ghahramani, J.R. Flanagan, Perspectives and problems in motor learning, Trends in Cognitive Sciences 5 (2001) 487–494.
[98] D.M. Wolpert, M. Kawato, Multiple paired forward and inverse models for motor control, Neural Networks 11 (1998) 1317–1329.
[99] D.M. Wolpert, R.C. Miall, Forward models for physiological motor control, Neural Networks 9 (1996) 1265–1279.