Mirror neurons and imitation: A computationally guided review

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

[a] JST-ICORP Computational Brain Project, Kyoto, Japan
[b] ATR, Computational Neuroscience Laboratories, 2-2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
[c] Computer Science, Neuroscience and USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA

Abstract

Neurophysiology reveals the properties of individual mirror neurons in the macaque while brain imaging reveals the presence of 'mirror systems' (not individual neurons) in the human. Current conceptual models attribute high level functions such as action understanding, imitation, and language to mirror neurons. However, only the first of these three functions is well-developed in monkeys. We thus distinguish current opinions (conceptual models) on mirror neuron function from more detailed computational models. We assess the strengths and weaknesses of current computational models in addressing the data and speculations on mirror neurons (macaque) and mirror systems (human). In particular, our mirror neuron system (MNS), mental state inference (MSI) and modular selection and identification for control (MOSAIC) models are analyzed in more detail. Conceptual models often overlook the computational requirements for posited functions, while too many computational models adopt the erroneous hypothesis that mirror neurons are interchangeable with imitation ability. Our meta-analysis underlines the gap between conceptual and computational models and points out the research effort required from both sides to reduce this gap.
© 2006 Published by Elsevier Ltd.

Keywords: Mirror neuron; Action understanding; Imitation; Language; Computational model

1. Introduction

Many neurons in the ventral premotor area F5 in macaque monkeys show activity in correlation with the grasp[1] type being executed (Rizzolatti, 1988). A subpopulation of these neurons, the mirror neurons (MNs), exhibit multi-modal properties responding to the observation of goal directed movements performed by another monkey or an experimenter (e.g. precision or power grasping) for grasps more or less congruent with those associated with the motor activity of the neuron (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). The same area includes auditory mirror neurons (Kohler et al., 2002) that respond not only to the view but also to the sound of actions with typical sounds (e.g. breaking a peanut, tearing paper). The actions associated with mirror neurons in the monkey seem to be transitive, i.e. to involve action upon an object and apply even to an object just recently hidden from view (e.g. Umilta et al., 2001).

It is not possible to find individual mirror neurons in humans since electrophysiology is only possible in very rare cases and at specific brain sites in humans. Therefore, one usually talks about a 'mirror region' or a 'mirror system' for grasping identified by brain imaging (PET, fMRI, MEG, etc.). Other regions of the brain may support mirror systems for other classes of actions. An increasing number of human brain mapping studies now refer to a mirror system (although not all are conclusive). Collectively these data indicate that action observation activates certain regions involved in the execution of actions of the same class. However, in contrast to monkeys, intransitive actions have also been shown to activate motor regions in humans. The existence of a (transitive and intransitive) mirror system in the human brain has also been supported by behavioral experiments illustrating the so-called 'motor interference' effect where observation of a movement degrades the performance of a concurrently executed incongruent movement (Brass, Bekkering, Wohlschlager, & Prinz, 2000; Kilner, Paulignan, & Blakemore, 2003; see also Sauser & Billard, this issue, for functional models addressing this phenomenon). Because of the overlapping neural substrate for action execution and observation in humans as well as other primates, many researchers have attributed high level cognitive

---

* Corresponding author. Address: Department of Cognitive Neuroscience, ATR, Computational Neuroscience Laboratories, 2-2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288, Japan. Tel.: +81 774 95 1215; fax: +81 774 95 1236.
E-mail address: erhan@atr.jp (E. Oztop).

[1] We restrict our discussion to hand-related neurons; F5 contains mouth-related neurons as well.

functions to MNs such as imitation (e.g. Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Miall, 2003), action understanding (e.g. Umilta et al., 2001), intention attribution (Iacoboni et al., 2005) and—on the finding of a mirror system for grasping in or near human Broca's area—(evolution of) language (Rizzolatti & Arbib, 1998).[2]

We stress that, although statements are often made about mirror *neurons* in humans, we have data only on what might be called mirror *systems* in humans—connected regions that are active in imaging studies both when the subject observes an action from some set and executes an action from that set, but not during an appropriate set of control tasks. Brain imaging results show that mirror *regions* in human may be associated with imitation and language (Carr et al., 2003; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Iacoboni et al., 1999; Skipper, Nusbaum, & Small, 2005), but there are no corresponding data on mirror *neurons*. Moreover, monkeys do not imitate (but see below) or learn language and so any account of the role of mirror neurons in imitation and language must include an account of the evolution of the human mirror system (Rizzolatti & Arbib, 1998) or at least the biological triggers that can unleash in monkeys a rudimentary imitation capability that goes beyond those they normally exhibit, though still being quite limited compared to those of humans (Kumashiro et al., 2003). We thus argue that imitation and language are not inherent in a macaque-like mirror system but instead depend on the embedding of circuitry homologous to that of the macaque in more extended systems within the human brain.

A general pitfall in conceptual modeling is that an innocent looking phrase thrown in the description may render the model implausible or trivial from a computational perspective, hiding the real difficulty of the problem. For example, terms like 'direct matching' and 'resonance' are used as if they were atomic processes that allow one to build hypotheses about higher cognitive functions of mirror neurons (Gallese, Keysers, & Rizzolatti, 2004; Rizzolatti, Fogassi, & Gallese, 2001). One must explain the cortical mechanisms which support the several processing stages that transform retinal stimulation caused by an action observation into the mirror neuron responses. Another issue is to clarify what is encoded by the mirror neuron activity. Is it the motor command, the meaning or the intention of the observed action? In an attempt to explain the multiplicity of functions attributed to mirror neurons, it has been recently speculated that different set of mirror neurons are involved in different aspects of the observed action (Rizzolatti, 2005).

We will review various 'conceptual models' that pay little attention to this crucial reservation, and then review several computational models in some detail: a learning architecture with parametric biases (Tani, Ito, & Sugita, 2004); a genetic algorithm model which develops networks for imitation while yielding mirror neurons as a by-product of the evolutionary process (Borenstein & Ruppin, 2005); the mirror neuron system (MNS) model that can learn to 'mirror' via self-observation of grasp actions (Oztop & Arbib, 2002) and is closely linked to macaque behavior and (somewhat more loosely) neurophysiology; and models which are not restricted in this fashion: the mental state inference (MSI) model that builds on the forward model hypothesis of mirror neurons (Oztop, Wolpert, & Kawato, 2005), the modular selection and identification for control (MOSAIC) model that utilizes multiple predictor–controller pairs (Haruno, Wolpert, & Kawato, 2001; Wolpert & Kawato, 1998), and the imitation architecture of Demiris and Hayes (2002) and Demiris and Johnson (2003).

2. Mirror neurons and action understanding

Mirror neurons, when initially discovered in macaques, were thought to be involved in action recognition (Fogassi et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996) though this laid the basis for later work ascribing a role in imitation to the human mirror system. Although the term 'action understanding' was often used, the exact meaning of 'understanding' as used here is not clear. It can range from 'act according to what you see', to 'infer the intentions/mental states leading to the observed action'. 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. In these experiments, he is given no opportunity to show by his behavior that he understands the action in either of the above senses.

Gallese and Goldman (1998) suggested 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.[3] Again, this involves considerable extrapolation beyond the available data. In particular, mind-reading might involve a quite separate mirror system for facial expression as much as a mirror system for manual actions. Although the suggestion has achieved some positive reception, no details have been provided on how this could be implemented as a computational model. (The MSI model does address this issue, see below).

In spite of the computational differences between recognition (and imitation) of facial gestures and recognition (imitation) of hand actions, many cognitive neuroscientists address them both under a 'generic mirror system'. The insula has been found to be a common face-emotion region for both production and understanding of facial gestures and emotions (Carr et al., 2003; Wicker et al., 2003). Although it is tempting to consequently take the insula as formed by instantiating

---

[2] Recent reviews of the mirror neuron and mirror system literature are provided by Buccino, Binkofski, and Riggio (2004), Fadiga and Craighero (2004) and Rizzolatti and Craighero (2004).

---

[3] Two predominant accounts of mindreading exist in the literature. 'Theory theory' asserts that mental states are represented as inferred conjectures of a naive theory whereas according to 'simulation theory', mental states of others are represented by representing their states as terms of one's own.

an F5-like mirror system for emotion processing, there are crucial differences in how these circuits develop in infancy which shed doubt on the idea of a generic neural mirror mechanism to unify social cognition (Gallese et al., 2004).

Manual actions can be compared visually and vocalizations can be compared auditorially—the commonality being the matching of an observed action with the output from an internal motor representation through a comparison in the same domain. Unlike hand actions, one's own facial gestures can only be seen with the help of a reflective material or otherwise must be inferred. In fact, infants may learn much about their own facial expressions from the propensity of caregivers to imitate the child. We argue that the learning mechanism of facial imitation is different from hand imitation, which involves learning via social interaction of the kinds we suggested elsewhere (Oztop et al., 2005):

You eat A→you have face expression X (visual).

I eat A→I feel disgust Y (internal state).

Therefore, X (visual) must be Y (feeling of disgust).

The elaboration of this mechanism in the brain is beyond the scope of this review. However, the important message here is that manual action understanding and facial emotion understanding pose different problems to the primate brain, which might have found solutions in different organizational principles.

3. Mirror neurons and imitation

Learning by imitation is an important part of human motor behavior, which requires complex set of mechanisms (Schaal, Ijspeert, & Billard, 2003). Wolpert, Doya, and Kawato (2003) underline some of those as (i) mapping the sensory variables into corresponding motor variables, (ii) compensation for the physical difference of the imitator from the demonstrator, and (iii) understanding the intention (goal) causing the observed movement. Many cognitive neuroscientists view imitation as mediated by mirror neurons in humans. Although this is a plausible hypothesis, we stress again that this is not within the normal repertoire of the macaque mirror system (but note the discussion below of Kumashiro et al., 2003) and so must—if true—rest on evolutionary developments in the mirror system. It should also be emphasized that there is a considerable amount of literature addressing imitation without explicit reference to mirror neurons (e.g. see Byrne & Russon, 1998).

Since most mirror neurons are found in motor areas it is reasonable to envision a motor control role for mirror neurons. One possibility is that these neurons implement an *internal model* for control. Current evidence suggests that the central nervous system uses internal models for movement planning, control, and learning (Kawato & Wolpert, 1998; Wolpert & Kawato, 1998). A *forward model* is one that predicts the sensory consequences of a motor command (Miall & Wolpert, 1996; Wolpert, Ghahramani, & Flanagan, 2001); while an *inverse model* transforms a desired sensory state into a motor command that can achieve it. The proposal of Arbib and Rizzolatti (1997) that mirror neurons may be involved in inverse modeling plays a central role in recent hypotheses about the neural mechanisms of imitation, which were accelerated by the increasing number of functional brain imaging studies.

One recent proposal is that the mirror neurons may provide not only an inverse model but a forward model of the body which can generate action candidates in the superior temporal sulcus (STS) where neurons have been found with selectivity for biological movement (e.g. of arms, whole body) (Carr et al., 2003; Iacoboni et al., 1999). The idea is that STS acts as a comparator that can be used within a search mechanism that finds the mirror neuron-projected action code that matches the observed action best that code is in turn used for subsequent imitation. It is further suggested that the STS-F5 circuit can be run in the reverse direction (inverse modeling) to map the observed action into motor codes (mirror activity) so that a rough motor representation of the observed act becomes available for imitation. According to this hypothesis, the F5-STS circuitry must be capable of producing detailed visual representation of the self-actions. From a computational point of view, if we accept that an observed act can be transformed into motor codes or if we accept the availability of an elaborate motor→visual forward model then imitation becomes trivial.

However, the above conceptual model limits imitation to actions already in the observer's repertoire. It is one thing to recognize a familiar action and quite another to see a novel action and consequently add it to one's repertoire. Another concern is that the mirror neurons found with electrophysiological recordings in monkeys are limited to *goal directed* actions. Indeed, behavioral studies show that chimpanzees cannot handle imitation tasks which do not involve any target objects (Myowa-Yamakoshi & Matsuzawa, 1999). In their efforts to release rudimentary imitation capability in monkeys (Kumashiro et al., 2003) used four tests based on objects (cotton-separation, knob-touching, latched box opening and removing the lid from a conical tube) and (with highly variable success) three tests based on facilitating use of a specific effector (tongue-protrusion model, hand-clench and thumb-extension) and two based on the movement of a hand relative to the body (hand to nose, hand clap, and hand to ear).

Miall (2003) suggested amending the conceptual model of Iacoboni et al. by including the cerebellum. He proposed that the forward and inverse computations required can be carried out by the cerebellum and PPC (posterior parietal cortex). The cerebellum has often been considered the likely candidate for (forward and inverse) internal models (Kawato & Gomi, 1992; Wolpert, Miall, & Kawato, 1998), but an alternative view (related to Miall's) is that cerebellar models act in parallel with models implemented in cerebral cortex, rather than replacing them (Arbib, Erdi, & Szentagothai, 1998). However, the computational problem is still there: how the retinal image is transformed to motor commands in a precise way (inverse problem), and how a precise visual description of one's own body is mentally produced given a motor command (forward problem). For inverse and forward models whose sensory data are limited to the visual domain, the problems are quite severe when the whole body is considered because one cannot completely observe all of one's own body. However,

observation of the hand in action is possible, enabling forward and inverse model learning. Presumably, the brain does not rely on visual data alone, but integrates it with proprioceptive cues. Although inverse learning is harder than forward learning, in general this does not pose a huge problem, and by using certain invariant representations in the visual domain, the learned internal models can be applied to other individuals' movements.[4]

4. Mirror neurons and language

Rizzolatti and Arbib (1998) built on evidence that macaque F5 is homologous to human Broca's area (an area often associated with speech production) and that human brain imaging reveals a mirror system for grasping in or around Broca's area to propose that brain mechanisms supporting language in humans evolved atop a primitive mirror neuron system similar to that found in monkeys. According to their *mirror system hypothesis*, human Broca's area's mirror system properties provide the evolutionary basis for language parity (the approximate meaning-equivalence of an utterance for both speaker and hearer). Arbib (2002, 2005) has expanded the hypothesis into seven evolutionary stages:

S1: Simple grasping,
S2: A mirror system for grasping,
S3: A simple imitation system for grasping,
S4: A complex imitation system for grasping,
S5: Protosign, a hand-based communication system,
S6: Protospeech, a vocalization-based communication system,
S7: Language, which required little or no biological evolution beyond S6, but which resulted from cultural evolution in *Homo sapiens*.

We do not have space here to comment on all the above steps but one quite distinctive feature of the analysis of stage S2 deserves attention. It asks the question 'why do mirror neurons exist?' and answers 'because mirror neurons are (originally) involved in motor control'. These neurons are located in the premotor cortex, interleaved with other motor neurons, where distal hand movements are controlled. Indeed, one may argue that the visual feedback required for manual dexterity—based on observation of the relation of hand to goal object—provided mechanisms that were exapted in primate and hominid evolution first for action recognition and then for imitation.

Although it is a daunting task to computationally realize this evolutionary model in its complete form (an attempt not without its critics—see the commentaries pro and con in Arbib (2005)), the transitions from one stage to the next can be potentially studied from a computational perspective.

_____

[4] In general it is not possible to infer the full dynamics from a kinematics observation, but we may assume some approximate solution that yields similar kinematics when applied by the observer.

5. Computational models involving MNs

We still lack a systematic neurophysiological study that correlates mirror neuron activity with the kinematics of the monkey or the demonstrator which would allow the computational modeler to test ideas about such correlations. Moreover, if (as we believe) the mirror properties of these neurons—as distinct from the conditions which make their acquisition possible—are not innate, then the study of the developmental course of these neurons and their function would test computational models of development that could help us better understand the functional role of MNs.

A general but wrong assumption in many computational studies of imitation is that mirror neurons are responsible for generating actions (and even sometimes that area F5 is composed only of mirror neurons). Indeed, F5 can be anatomically subdivided into two distinct regions, one containing the mirror neurons, and one containing canonical neurons. The latter are like mirror neurons in their motor properties, but do not respond to action observation. Muscimol injections (muscimol causes reversible neural inactivation) to the part of area F5 that includes mirror neurons do not impair grasping and control ability, but causes only a slowing down of the action (Fogassi et al., 2001).[5] However, when the area that includes the canonical neurons is the target of injection, the hand shaping preceding grasping is impaired and the hand posture is not appropriate for the object size and shape (Fogassi et al., 2001).

In the following sections, we will review various computational studies that relate (often at some conceptual remove) to mirror neurons. Unfortunately, most of the modeling is targeted at imitation. Only one, the MNS model of Oztop and Arbib (2002) directly claims to be a model for mirror neurons (although it does not provide computational modules for motor control).

6. A dynamical system approach

The first model we review, due to Jun Tani et al. (2004), is aimed at learning, imitation and autonomous behavior generation. The proposed network is a generative learning architecture called recurrent neural network with parametric biases (RNNPB). In this architecture, the spatio-temporal patterns are associated with so-called parametric bias vectors (PB). RNNPB self-organizes the mapping between PBs and the spatio-temporal patterns (behaviors) during the learning phase. From a functional point of view the goal of RNNPB is similar to dynamical movement primitive learning (Ijspeert, Nakanishi, & Schaal, 2003; Schaal, Peters, Nakanishi, & Ijspeert, 2004) in that the behaviors are learned as dynamical systems. However, in the dynamical movement primitive approach, dynamical systems are not constructed from scratch as

_____

[5] This may at first appear inconsistent with the view that mirror neurons may assist the feedback control for dexterous movements, but note that the muscimol studies were only carried out for highly familiar grasps whose successful completion would require little if any visual feedback.

in RNNPB; but rather (core) primitive dynamical systems are adapted to match the demonstrated movements using local learning techniques (Ijspeert et al., 2003; Schaal et al., 2004). In this article, we focus on RNNPB as the representative of dynamical system approaches since the authors have already hinted that RNNPB captures some properties of the mirror neurons (Tani et al., 2004). The RNNPB has three operational modes; we review each of them starting from the learning mode.

6.1. Learning mode

The learning is performed in an off-line fashion by providing the sensory-motor training stimuli (e.g. two trajectories—one for the position of a moving hand and the other for the joint angles of the arm) for each behavior in the training set. The goal of the training is twofold (see Fig. 1): (1) to adapt weight sets w and W such that the network becomes a time series predictor for the sensory-motor stimuli, and (2) to create PB vectors for each training behavior. Both adaptations are based on the *prediction error*; the weights w and W are adapted by back-propagation over all the training patterns as in a usual recurrent neural network. The PB vectors, however, are updated separately for each training pattern for reducing the prediction error. Furthermore, the modulation of PB vectors are kept slow to obtain a fixed PB vector for each learnt behavior.

6.2. Action generation mode

After learning, the model represents a set of behaviors as dynamical systems tagged by the PB vectors created during the learning phase. The behaviors are generated via the associated PB vectors. Given a fixed PB vector, the network autonomously produces a sensory–motor stream corresponding to the behavior associated with the PB vector (see Fig. 2). Note that the behavior generation mode of RNNPB requires the sensory–motor prediction to be fed back into the sensory–motor input as shown in Fig. 2.
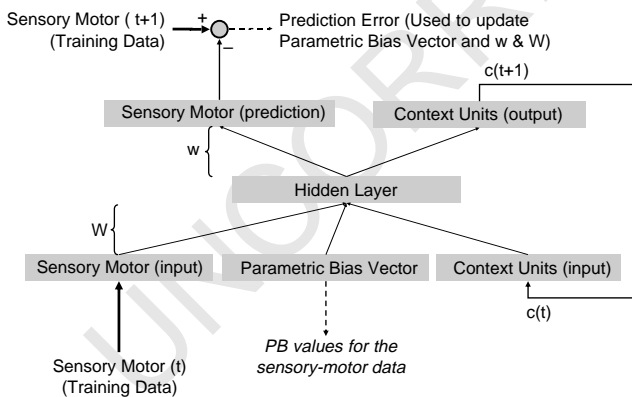

Fig. 1. The RNNPB network in learning phase. The weights W and w are adapted so as to reduce the prediction error over all the behaviors to be learnt (temporal sensory–motor patterns). The PB vectors are also updated to reduce the prediction error, albeit separately for each behavior and at a slower rate to ensure representative PB values for each behavior.
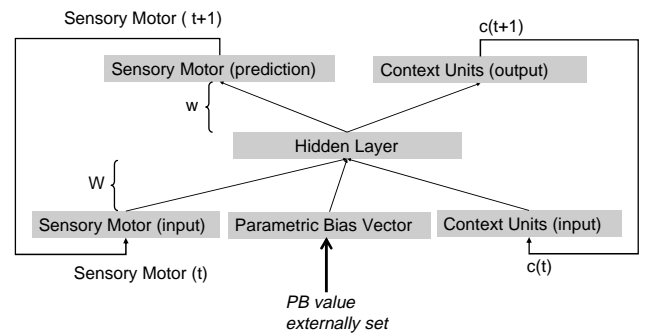

Fig. 2. The RNNPB in behavior generation mode. Given a fixed PB vector (thick arrow) the network produces the corresponding stored sensory–motor stream.

6.3. Action recognition mode

The task of the network in this mode is to observe an ongoing behavior (sensory data) and compute a PB vector that is associated with a behavior that matches the observed one as much as possible. The arrival of a *sensory input* generates a *prediction* of the next sensory stimulus at the output layer. Then the *actual next sensory input* is compared with this prediction creating a *prediction error* (see Fig. 3). The prediction error is back-propagated to the PB vectors (i.e. PB vectors are updated such that the prediction error is reduced). The actual computation of the PB vectors is performed using the so-called regression window of the past steps so that the change of PB vectors can be constrained to be smooth over the window. The readers are referred to Tani et al. (2004) for further details of this step. If sensory input matches one of the learned behaviors, the PB vectors tend to converge to the values determined during learning (Tani et al., 2004). Note that in this mode, the feedback from sensory–motor output to sensory-motor input is restricted only to the motor component.

6.4. Relation to MNs and imitation

The model has been shown to allow a humanoid robot to imitate and learn actions via demonstration. The logical link
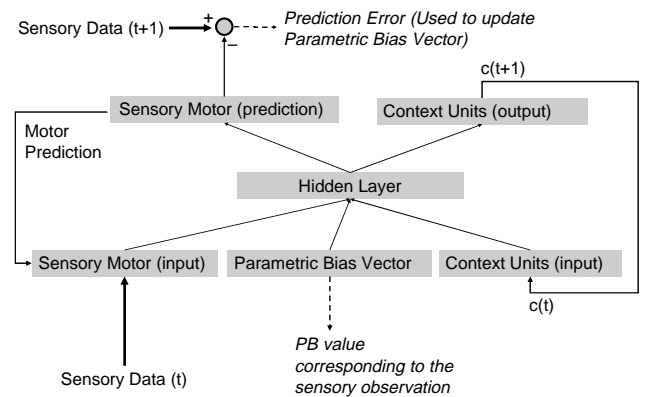

Fig. 3. The RNNPB in behavior recognition mode. In the recognition mode the sensory input is obtained from external observation (thick arrows). The feedback from sensory–motor output to sensory–motor input is restricted only to the motor component. The prediction error is used to compute the parametric bias vector corresponding to the incoming sensory data.

to mirror neurons comes from the fact that the system works as both a behavior recognizer and generator after learning. PB units are tightly linked with the behavior being executed or observed. During execution, a fixed PB vector selects one of the stored motor patterns. For recognition PB unit outputs iteratively converge to the action observed. Although mirror neurons do not determine the action to be executed in monkeys (Fogassi et al., 2001), the firing patterns of mirror neurons are correlated with the action being executed. Thus, PB vector units may be considered analogous to mirror neurons. Ito and Tani (2004) suggest that the PB units' activities should be under control of a higher mechanism to avoid unwanted imitation, such as for dangerous movements. One prediction, or rather a question posed to neurophysiology, is what happens when a dangerous movement is observed by a monkey. Although it is known that (initially) F5 mirror neurons do not respond to unfamiliar actions, no data on the parietal mirror neurons exists to rule out this possibility. However, we again emphasize that mirror response does not automatically involve movement imitation and so it is unlikely that the monkey mirror neuron system is 'an inhibited imitation system'; rather imitation ability must have been developed on top of the mirror neuron system along the course of primate evolution.

7. Motor learning and imitation: a modular architecture

The RNNBP of the previous section represents multiple behaviors as a distributed code. Demiris et al. (2002, 2003) chose the opposite approach representing each behavior as a separate module in their proposed imitation system. Following the organization of the MOSAIC model (Wolpert & Kawato, 1998) (see Section 12), the key structure of the proposed architecture is formed by a battery of behaviors (modules) paired with forward models, where each behavior module receives information about the current state (and possibly the target goal), and outputs the motor commands that is necessary to achieve the associated behavior (see Fig. 4). A forward model receives output of the paired behavior module and estimates the next state which is fed back to the behavior module for parameter adjustments. A behavior is similar to an 'inverse model', although inverse models do not usually utilize
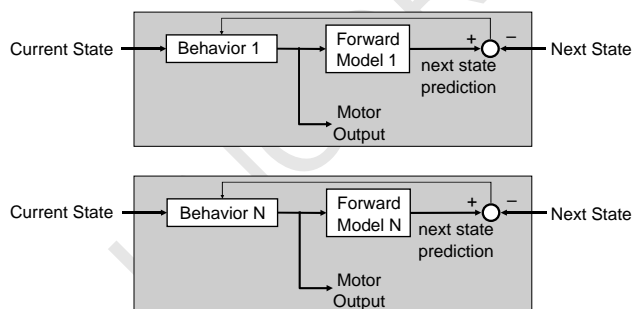

Fig. 4. The imitation architecture proposed by Demiris et al. (2002, 2003) is composed of a set of paired behavior and forward models. During imitation mode, the comparison of the predicted next state (of the demonstrator) with the actual observed state gives an indication of which behavior module should be active for the correct imitation of the observed action.

feedback, but output commands in a feed-forward manner. However, the boundary between behavior and inverse model is not a rigid one (Demiris & Hayes, 2002).

The architecture implements imitation by assuming that the demonstrator's current state (e.g. joint angles of a robot) is available to it. When the demonstrator executes a behavior, the perceived states are fed into the imitator's available behavior modules in parallel which generate motor commands that are sent to the forward models. The forward models predict the next state based on the incoming motor commands, which are then compared with the actual demonstrator's state at the next time step. The error signal resulting from this comparison is used to derive a confidence value for each behavior (module). The behavior with the highest confidence value (i.e. the one that best matches the demonstrator's behavior) is selected for imitation. When an observed behavior is not in the existing repertoire, none of the existing behaviors reach a high confidence value, thus indicating that a new behavior should be added to the existing behavior set. This is achieved by extracting representative postures while the unknown behavior is demonstrated, and constructing a behavior module (e.g. a PID controller) to go through the representative postures extracted. This computational procedure to estimate the other agent's behavioral module by simulating one's own forward model and controller is essentially identical to the proposal by Doya, Katagiri, Wolpert, and Kawato (2000). Demiris and Simmons (this issue) describe a hierarchical architecture that employs similar principles at its core. Although at a conceptual level, this architecture has strong parallels with the MOSAIC model (Haruno et al., 2001; Wolpert & Kawato, 1998), MOSAIC takes *learning and control* as the core focus by providing explicit learning mechanisms (see Section 12).

7.1. Relation to MNs and imitation

The architecture can be related to mirror neurons because the behavior modules are active during both movement generation and observation. However, all the modules are run in parallel in the proposed architecture, so it is more reasonable to take the confidence values as the mirror neuron responses. Demiris et al. (2002, 2003) arrived at several predictions about mirror neurons albeit considering 'imitation-ability' and 'mirror neuron activity' interchangeable, while we think they must be analyzed separately. One interesting prediction which has also been predicted by the MNS model (Oztop & Arbib, 2002) (see Section 10) is the following. 'A mirror neuron which is active during the demonstration of an action should not be active (or possibly be less active) if the demonstration is done at speeds unattainable by the monkey'. A further prediction states that 'mirror neurons that remain active for a period of time after the end of the demonstration are encoding more complex sequences that incorporate the demonstration as their first part' (Demiris & Hayes, 2002). The other predictions—implied by the structure of the architecture—are 'the existence of other goal directed mirror neurons and the trainability of new mirror neurons'.

8. An evolutionary approach

The evolutionary algorithms that Borenstein and Ruppin (2005) used to explain mirror neurons and imitation are quite different from the previous approaches. Evolutionary algorithms incorporate aspects of natural selection (survival of the fittest) to solve an optimization problem. An evolutionary algorithm maintains a population of structures ('individuals') that evolves according to rules of selection, recombination, mutation and survival. A shared 'world' determines the fitness or performance of each individual and identifies the optimization problem. Each 'generation' is composed of fitter individuals and their variants, while fewer not-fit individuals and their variants are allowed to reproduce their traits. After many generations one expects to find a set of high performing individuals which represent close-to-optimal solutions to the original problem. Within this framework, Borenstein and Ruppin (2005) defined individuals as simple neuro-controllers that could sense the state of the world and the action of a *teaching agent* (inputs) and generate actions (outputs). Individuals generated output with a simple 1-hidden-layer feedforward neural network. The fitness of an individual was defined as a random mapping from world state to actions, which was kept fixed for the lifetime of an individual. The evolutionary encoding (genes) determined the properties of individual network connections (synapses): type of learning, initial strength, either inhibitory or excitatory type, and rate of plasticity.

8.1. Relation to MNs and imitation

The simulation with populations of 200 individuals evolving for 2000 generations showed interesting results. The best individuals developed neural controllers that could learn to imitate the teacher. Furthermore, the analysis of the units in the hidden layer of these neuro-controllers revealed units which were active both when observing the teacher and when executing the correct action, although not all the actions were mirrored. The conclusion drawn was that there is an 'essential link between the ability to imitate and a mirror system'. Despite the interest of the demonstration that evolving neural circuitry to imitate yields something like mirror neurons as a byproduct, we note again that even though monkeys have mirror neurons they are not natural imitators. Thus, the 'evolution' of mirror systems on the basis of 'evolutionary pressure' to imitate does not seem to capture the time course of primate evolution.

9. Associative memory hypothesis of mirror neurons

In this section, we avoid choosing a single architecture for extensive review since the core mechanism employed in all the candidate models relies on a very simple principle deriving from the classical view of Hebbian synaptic plasticity in the cerebral cortex. Implementation of this view results in connectionist architectures referred as associative or content addressable memories (Hassoun, 1993), to which the models
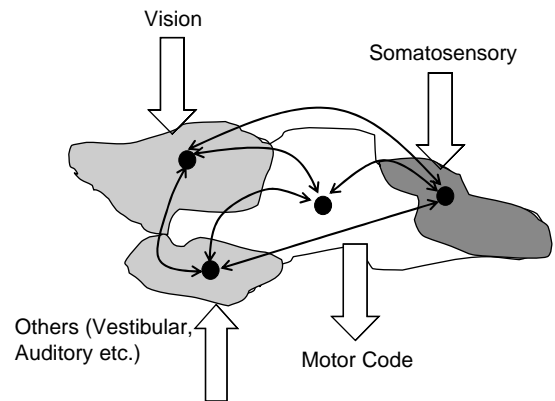

Fig. 5. A generic associative memory for an agent (a robot or an organism). When an agent generates a movement using motor code, the sensed stimuli are associated with this code. At a later time a partial representation of associated stimuli (e.g. vision) can be used to retrieve the whole (including the motor code). The connectivity among the units representing different modalities could be full or sparse.

reviewed in this section more or less conform. The crucial feature of an associative memory is that a partial representation of a stored pattern is sufficient to reconstruct the whole. In general a neural network which does not distinguish input and output channels can be considered as an associative memory (with possibly hidden units). Fig. 5 schematizes a possible association that can be established when a biological or an artificial agent acts. The association can take place among the *motor code*, the *somatosensory*, *vestibular*, *auditory* and *visual* stimuli sensed when the movement takes place with the execution of the motor code. If we hypothesize that mirror neurons are part of a similar mechanism then the mirror neuron responses could be explained: when the organism generates motor commands the representation of this command and the sensed (somatosensory, visual and auditory) effects of the command are associated within the mirror neuron system. Then at a later time when the system is presented with a stimulus that partially matches one of the stored patterns (i.e. vision or audition of an action alone) the associated motor command representation is retrieved automatically. This representation can be used (with additional circuitry) to mimic the observed movement. This line of thought has been explored through robotic implementations of imitation using a range of associative memory architectures. Elshaw, Weber, Zochios, and Wermter (2004) implemented an associator network based on the Helmholtz machine (Dayan, Hinton, Neal, & Zemel, 1995) where the motor action codes were associated with vision and language representations. The learned association enabled neurons of the hidden layer of the network to behave as mirror neurons; the hidden units could become active with one of motor, vision or language inputs. Kuniyoshi, Yorozu, Inaba, and Inoue (2003) used a spatio-temporal associative memory called the 'non-monotone neural net' (Morita, 1996) to associate self generated arm movements of a robot with the local visual flow generated. Billard and Mataric (2001) used the DRAMA architecture (Billard & Hayes, 1999), which is a time-delay recurrent neural network with Hebbian update

dynamics at the core of their biologically inspired imitation architecture. Oztop, Chaminade, Cheng, and Kawato (2005) used an extension of a Hopfield net utilizing product terms to implement a hand posture imitation system using a robotic hand.

9.1. Relation to MNs and imitation

In spite of the differences in the implementation, the common property among the aforementioned associative memory models is the *multi-modal activation* of the associative memory/network units. Thus, when these models are considered as models of mirror neurons (note that not all models claim to be models for mirror neurons, as the main focus is on imitation) then the explanation of the existence of mirror neurons becomes phenomenological rather than functional (see Section 13). For example, the models of Kuniyoshi et al. (2003) and Oztop et al. (2005) use self-observation as the principle for bootstrapping imitation and formation of units that respond to self-actions and observations of others. We refer to this type approach as ASSOC so that we can collectively refer to these models in Section 13, where we propose a taxonomy based on the modeling methodology.

10. Mirror neuron system (MNS) model: a developmental view

The models reviewed so far related to the mirror neurons indirectly, through imitation. Here, we present a computational model with anatomically justified connectivity, which directly explores how MNs develop during infancy. It is quite unlikely that the MNs are innate because mirror neurons have been observed for tearing paper, for instance (Kohler et al., 2002). With this observation, the MNS model (Oztop & Arbib, 2002) takes a developmental point of view and explains how the mirror neurons are developed during infancy. The main hypothesis of the model is that the temporal profile of the features an infant experiences during self-executed grasps provides the training stimuli for the mirror neuron system to develop.[6] Thus, developmentally, grasp learning precedes initial mirror neuron formation.[7] Although MNS proposes that MNs are initially evolved to support motor control, it does not provide computational mechanisms showing this.

10.1. The Model

MNS is a systems level model of the (monkey) mirror neuron system for grasping. The computational focus of the

model is the development of mirror neurons by self-observation; the motor production component of the system is assumed to be in place and not modeled using neural modules. The schemas[8] (Arbib, 1981) of the model are implemented with different level of granularity. Conceptually those schemas correspond to brain regions as follows (see Fig. 6). The inferior premotor cortex plays a crucial role when the monkey itself reaches for an object. Within the inferior premotor cortex area F4 is located more caudally than area F5, and appears to be primarily involved in the control of proximal movements (Gentilucci et al., 1988), whereas the neurons of F5 are involved in distal control (Rizzolatti et al., 1988). Areas IT (inferotemporal cortex) and cIPS (caudal intraparietal sulcus) provide visual input concerning the nature of the observed object and the position and orientation of the object's surfaces, respectively, to AIP. The job of AIP is to extract the affordances the object offers for grasping. By affordance we mean the object properties that are relevant for grasping such as the width, height and orientation. The upper diagonal in Fig. 6 corresponds to the basic pathway AIP→F5 canonical→M1 (primary motor cortex) for distal (reach) control. The lower right diagonal (MIP/LIP/VIP→F4) of Fig. 6 provides the proximal (reach) control portion of the MNS model. The remaining modules of Fig. 6 constitute the sensory processing (STS and area 7a) and the core mirror circuit (F5 mirror and area 7b).

Mirror neurons do not fire when the monkey sees the hand movement or the object in isolation; the sight of the hand moving appropriately to grasp or manipulate a seen (or recently seen) object is necessary for the mirror neurons tuned to the given action to fire (Umiltá et al., 2001). This requires schemas for the recognition of the shape of the hand and the analysis of its motion (performed by STS in the model), and for the analysis of the hand-object relation (Fig. 7a). The information gathered at STS and areas 7a are captured in the 'hand state' at any instant during movement observation and serves as an input to the core mirror circuit (F5 mirror and area Fig. 7b). Although visual feedback control was not built into MNS, the hand state components track the position of hand and fingers relative to the object's affordance (see Oztop and Arbib (2002) for the full definition of the hand state) and can thus be used in monitoring the successful progress of a grasping action supporting motor control. The crucial point is that the information provided by the hand state allows action recognition because relations encoded in the hand state form an invariant of the action regardless of the agent of the action. This allows self-observation to train a system that can be used for detecting the actions of others and recognizing them as one of the actions of the self.

During training, the motor code represented by active F5 canonical neurons was used as the training signal for the core mirror circuit to enable mirror neurons to learn which

---

[6] Only grasp related visual mirror neurons were addressed. A subsequent study (Bonaiuto et al., 2005) has introduced a recurrent network learning architecture that not only reproduces key results of Oztop and Arbib (2002) but also addresses the data of Umiltá et al. (2001) on grasping of recently obscured objects and of Kohler et al. (2002) on audiovisual mirror neurons.

[7] Note again that monkeys have a mirror system but do not imitate. It is thus a separate question to ask "How, in primates that do imitate, does the imitation system build (both structurally and temporally) on the mirror system?".

[8] A schema refers to a functional unit that can be instantiated as a modular unit, or as a mode of operation of a network of modules, to fulfill a desired input/output requirement (Arbib, 1981; Arbib et al., 1998).
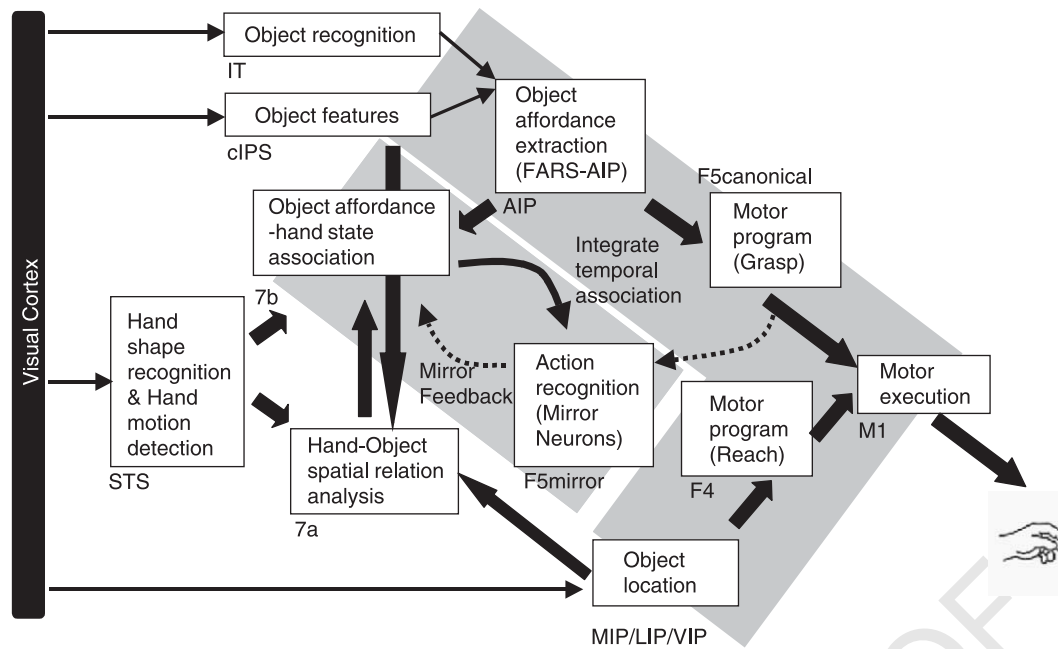
Fig. 6. A schematic view of the mirror neuron system (MNS) model. The MNS model learning mechanisms and simulations focus on the core mirror circuit marked by the central diagonal rectangle (7b and F5 mirror), see text for details (Oztop & Arbib, 2002; reproduced with kind permission of Springer Science and Business Media).

hand-object trajectories corresponded to the canonically encoded grasps. We reiterate that the input to the F5 mirror neurons is not the visual stimuli as created by the hand and the object in the visual field but the 'hand state trajectory' (trajectory of the relation of the hand and the object) extracted from these stimuli. Thus, training tunes the F5 mirror neurons to respond to hand-object relational trajectories independent of the owner of the action ('self' or 'other').

10.2. Relation to MNs

The focus of the simulations was the 7b-F5 complex (core mirror circuit). The input and outputs of this circuit was computed using various schemas providing a context to analyze the circuit. The core mirror circuit was implemented as a feedforward neural network (1-hidden layer back-propagation network with sigmoidal activation units; hidden layer: area 7b; output layer: F5 mirror) responding to increasingly long initial segments of the hand-state trajectory. The network could be trained to recognize the grasp type from the hand state trajectory, with correct classification often being achieved well before the hand reached the object. For the preprocessing and training details the reader is referred to Oztop and Arbib (2002).

Despite the use of a non-physiological neural network, simulations with the model generated a range of predictions about mirror neurons that suggest new neurophysiological experiments. Notice that the trained network responded not only to hand state trajectories from the training set, but also showed interesting responses to novel grasping modes. For example Fig. 7 shows one prediction of the MNS model. An ambiguous precision pinch grasp activates multiple neurons

(power and precision grasp responsive neurons) during the early portion of the movement observation. Only later does the activity of the precision pinch neuron dominate and the power grasp neuron's activity diminish.

Other predictions were derived from the spatial perturbation experiment where the hand did not reach the goal (i.e. a 'fake' grasp), and the altered kinematics experiment where the hand moved with constant velocity profile. The former case showed a non-sharp decrease in the mirror neuron activity while the latter showed a sharp decrease. The reader is referred to Oztop and Arbib (2002) for the details and other simulation experiments.

Recently, Bonaiuto, Rosta, and Arbib (2005) developed the MNS2 model, a new version of the MNS model of action recognition learning by mirror neurons of the macaque brain, using a recurrent architecture that is biologically more plausible than that of the original model. Moreover, MNS2 extends the capacity of the model to address data on audio–visual mirror neurons (Kohler et al., 2002) and on response of mirror neurons when the target object was recently visible but is currently hidden (Umiltá et al., 2001).

11. The mental state inference (MSI) model: forward model hypothesis for MNs

The anatomical location (i.e. premotor cortex) and motor response of mirror neurons during grasping suggest that the fundamental function of mirror neurons may be rooted in grasp control. The higher cognitive functions of mirror neurons, then should be seen as a later utilization of this system, augmented with additional neural circuits. Although MNS of the previous
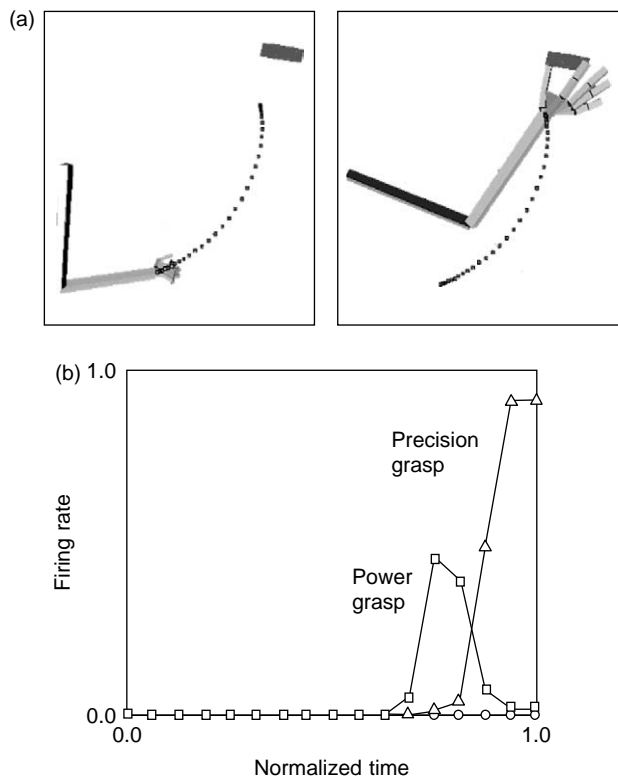
Fig. 7. Power and precision grasp resolution. (a) The left panel shows the initial configuration of the hand while the right panel shows the final configuration of the hand, with circles showing positions of the wrist in consecutive frames of the trajectory. (b) The distinctive feature of this trajectory is that the hand initially opens wide to accommodate the length of the object, but then thumb and forefinger move into position for a precision grip. Even though the model had been trained only on precision grips and power grips separately, its response to this input reflects the ambiguities of this novel trajectory—the curves for power and precision cross towards the end of the action, showing the resolution of the initial ambiguity by the network. (Oztop & Arbib, 2002, reproduced with kind permission of Springer Science and Business Media).

section adopted this view, it did not model the motor component, which is addressed by the MSI model.

11.1. Visual feedback control of grasping and the forward model hypothesis for the mirror neurons

The mental state inference (MSI) model builds upon a visual feedback circuit involving the parietal and motor cortices, with a predictive role assigned to mirror neurons in area F5. For understanding others' intentions, this circuit is extended into a mental state inference mechanism (Oztop et al., 2005). The global functioning of the model for visual feedback control proceeds as follows. The parietal cortex extracts visual features relevant to the control of a particular goal-directed action (X, the control variable) and relays this information to the premotor cortex. The premotor cortex computes the motor signals to match the parietal cortex output (X) to the desired neural code (X_{des}) relayed by prefrontal cortex. The 'desired change' generated by the premotor cortex is relayed to dynamics related motor centers for execution (Fig. 8, upper panel). The F5 mirror neurons implement a forward prediction

circuit (forward model) estimating the sensory consequences of F5 motor output related to manipulation, thus compensating for the sensory delays involved in a visual feedback circuit. This is in contrast to the generally suggested idea that mirror neurons serve solely to retrieve an action representation that matches the observed movement. During observation mode, these F5 mirror neurons are used to create motor imagery or mental simulation of the movement for mental state inference (see below). Although MSI does not specify the region within parietal cortex that performs control variable computation, recent findings suggest that a more precise delineation is possible. Experiments with macaque monkeys indicate that parietal area PF may be involved in monitoring the relation of the hand with respect to an object during grasping. Some of the PF neurons that do not respond to vision of objects become active when the monkey (without any arm movement) watches movies of moving hands (of the experimenter or the monkey) for manipulation, suggesting that the neural responses may reflect the visual feedback during observed hand movements (Murata, 2005). It is also possible that a part of AIP may be involved in monitoring grasping as shown by transcranial magnetic stimulation (TMS) with humans (Tunik, Frey, & Grafton, 2005). As in the MNS model, area F5 (canonical) is involved in converting the parietal output (PF/AIP) into motor signals, which are used by primary motor cortex and spinal cord for actual muscle activation. In other words, area F5 non-mirror neurons implement a control policy (assumed to be learned earlier) to reduce the error represented by area PF/AIP output.

11.2. Mental state inference

The ability to predict enables the feedback circuit of Fig. 8 (upper panel) to be extended into a system for inferring the intentions of others based on the kinematics of goal directed actions (see Fig. 8 lower panel). In fact, the full MSI model involves a 'mental simulation loop' that is built around a forward model (Blakemore & Decety, 2001; Wolpert & Kawato, 1998), which in turn is used by a 'mental state inference loop' to estimate the intentions of others. The MSI model is described for generic goal-directed actions, however here we look at the model in relation to a tool grasping framework where two agents can each grasp a virtual hammer with different intentions (holding, nailing or prying a nail). Depending on the planned subsequent use of a hammer, grasping requires differential alignment of the hand and the thumb. Thus, the kinematics of the action provides information about the intention of the actor. For this task, the mental state was modeled as the intention in grasping the hammer. Within this framework an observer 'guesses' the target object (one of various objects in the demonstrator's workspace) and the type of grasp and produce an appropriate F5 motor signal that is inhibited for actual muscle activation but used by the forward model (MNs) (see Fig. 8 lower panel). With the sensory outcome predicted by the MNs, the movement can be simulated as if it were executed in an online feedback mode. The match of the simulated sensations of a simulated
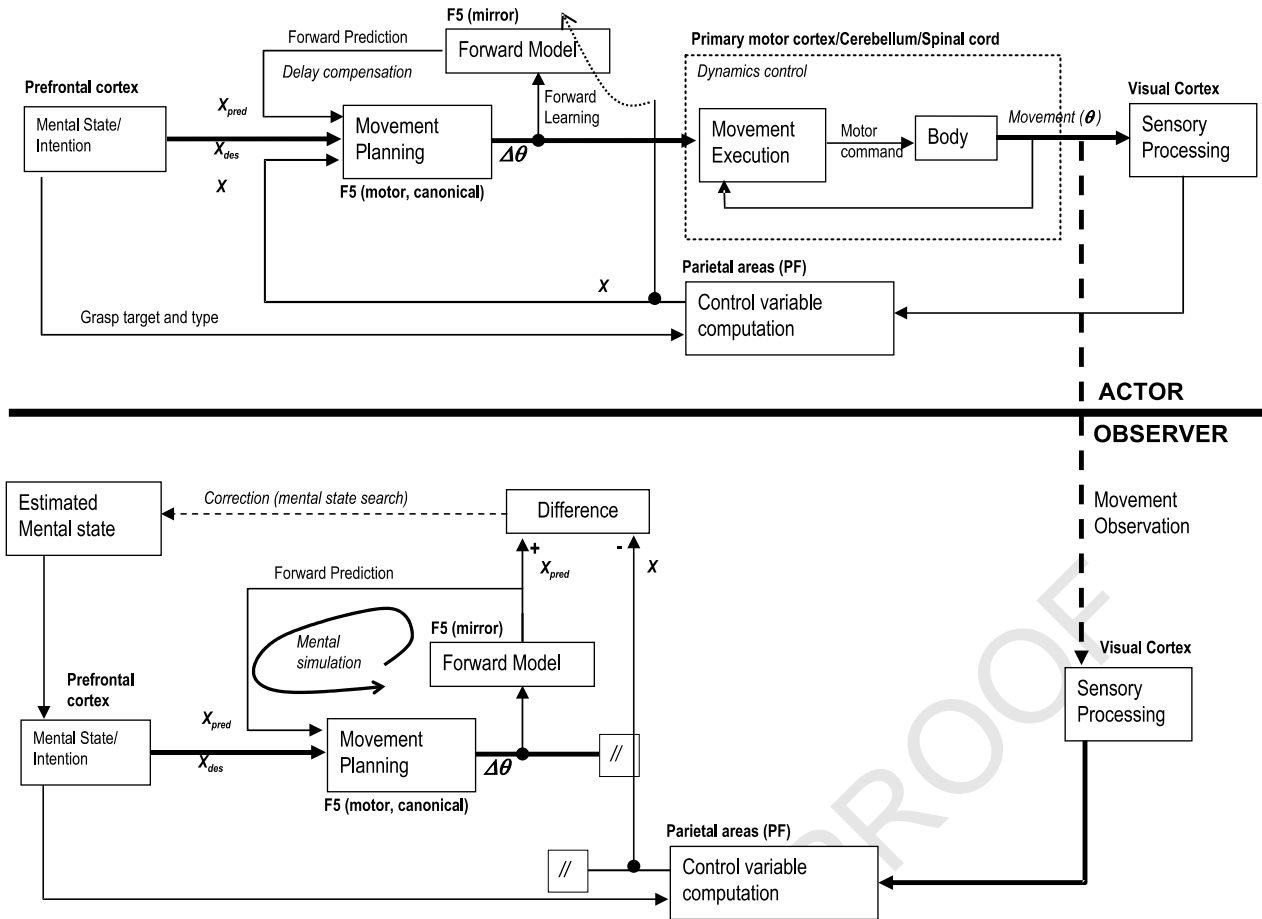
Fig. 8. Upper panel: the MSI model is based on the illustrated visual feedback control organization. Lower panel: observer's mental state inference mechanism. Mental simulation of movement is mediated by utilizing the sensory prediction from the forward model and by inhibiting motor output. The difference module computes the difference between the visual control parameters of the simulated movement and the observed movement. The mental state estimate indicates the current guess of the observer about the mental state of the actor. The difference output is used to update the estimate or to select the best mental state. (Adapted from Oztop et al., 2005).

movement with the sensation of observed movement will then signal the correctness of the guess. The simulated mental sensations and actual perception of movement is compared in a mental state search mechanism. If the observer model 'knows' the possible mental states in terms of discrete items an exhaustive search in the mental state space can be performed. However, if the mental state space is not discrete then a gradient based search strategy must be applied. The mental state correction (i.e. the gradient) requires the parietal output (PF) based errors (the Difference box in Fig. 8) to be converted into 'mental state' space adjustments, for which a stochastic gradient search can be applied (see Oztop et al. (2005) for the details).

11.3. Relation to MNs and Imitation

A tool-use experiment was set-up in a kinematics simulation where two agents could grasp a virtual hammer. The visual parameters used to implement the feedback servo for grasping (i.e. normalized distance and the orientation difference—see Fig. 9) were object centered and provided generalization regardless of the owner of the action (self vs. other). The time varying (mental simulation)×(observation) matrix shown in Fig. 10 represents the dynamics of an agent observing an actor performing tasks of holding, nailing and prying (rows of the matrix). Each column of the matrix represents the belief of the observer as to whether the other is holding, nailing or prying.
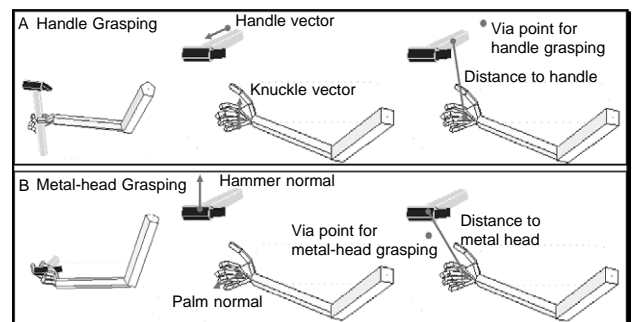

Fig. 9. (A) The features used for nailing task (orientation and normalized distance) are depicted in the right two arm drawings. The path of the hand is constrained with appropriate via-points avoiding collision. The arm drawing on the left shows an example of a handle grasping for driving a nail. The prying task is same as A except that the Handle vector points towards the opposite direction (not shown). (B) The features extracted for metal-head grasping is depicted (conventions are the same as the upper panel). (Adapted from Oztop et al., 2005).
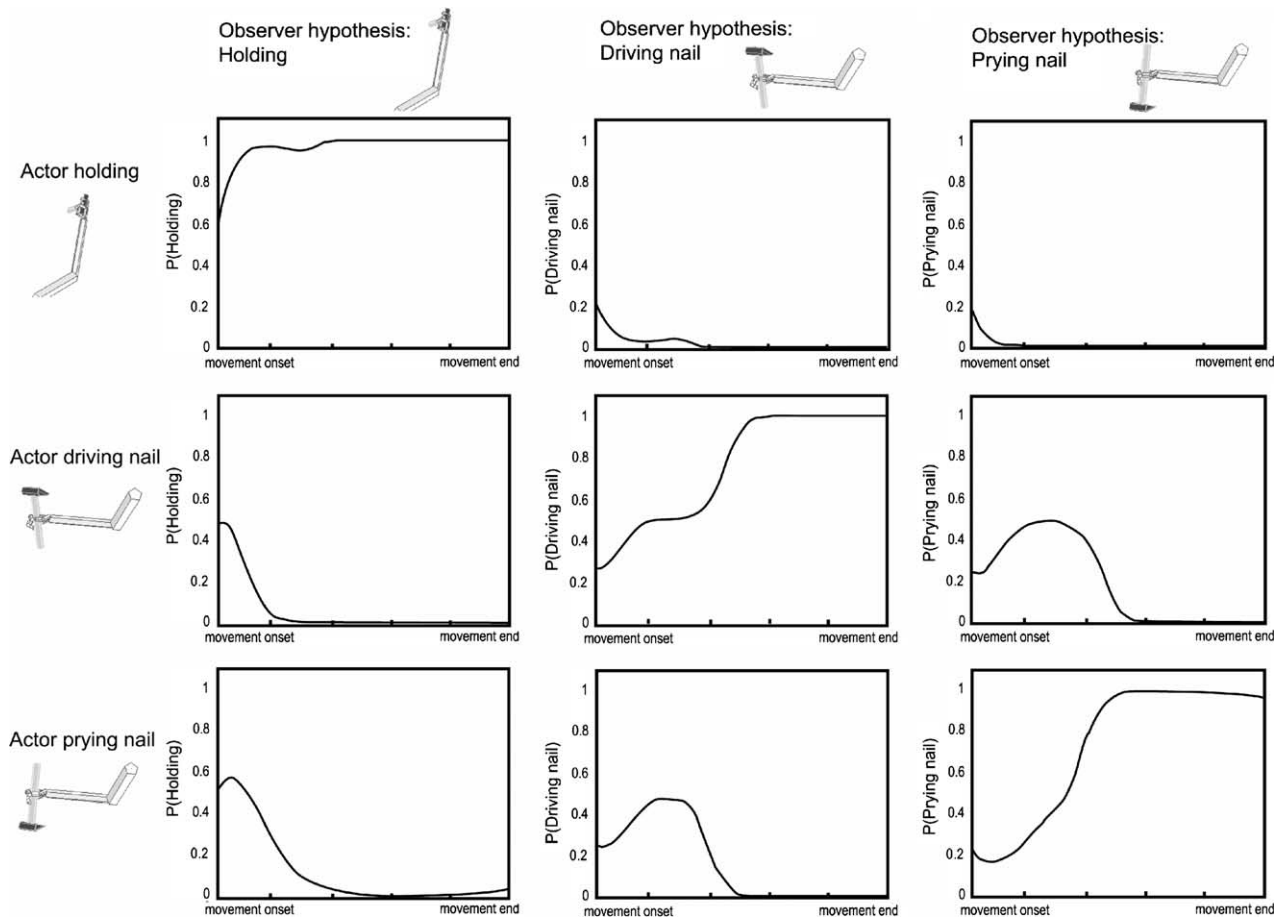
Fig. 10. The degree of similarity between visually extracted control variables and control variables obtained by mental simulation can be used to infer the intention of an actor. Each subplot shows the probability that the observed movement (rows) is the same as the mentally simulated one (columns). The horizontal axis represents the simulation time from movement start to end. The control variables extracted for the comparison are based on the mentally simulated movement. Thus, for example, the first column inferences require the control parameters for holding (normalized distance to metal head and the angle between the palm normal and hammer plane). The convergence to unity of the belief curves on the main diagonal indicates correct mental state inference. (Adapted from Oztop et al., 2005).

Each cell shows the degree of similarity between mentally simulated movement and the observed movement from movement onset to movement end (as a belief or probability). The observer can infer the mental state of the actor before the midpoint of the observed movements as evidenced by the convergence of belief curves to unity along the diagonal plots. Thus, although not the focus of the study, the MSI model offers a basic imitation ability that is based on reproducing the inferred intention (mental state) of an observer. However, the actions that can be imitated are thus limited to the ones in the existing repertoire and may not respect the full details of the observed act. With MSI the dual activation of MNs (forward model) is explained by the automatic engagement of mental state inference during an action observation, and by the forward prediction task undertaken by the MNs for motor control during action execution.

12. Modular selection and identification for control (MOSAIC) model

The MOSAIC model (Haruno et al., 2001; Wolpert & Kawato, 1998) was introduced initially for motor control, providing mechanisms for decentralized automatic module selection so as to achieve best control for the current task. In this sense, compared to the earlier models surveyed, MOSAIC is a sophisticated motor control architecture. The key ingredients of MOSAIC are modularity and the distributed cooperation and competition of the internal models. The basic functional units of the model are multiple predictor–controller (forward–inverse model) pairs where each pair competes to contribute to the overall control (cf. Jacobs, Jordan, Nowlan, and Hinton (1991), where the emphasis is on selection of a single processor). The controllers with better predicting forward models (i.e. with higher responsibility signals) become more influential in the overall control (Fig. 11). The responsibility signals are computed with the normalization step shown in Fig. 11 based on the prediction errors of the forward models via the softmax function (see Table 1, last row). The responsibility signals are constrained to be between 0 and 1, and add up to 1, so that that they can be considered as probabilities indicating the likelihood of the controllers being effective for the current task.
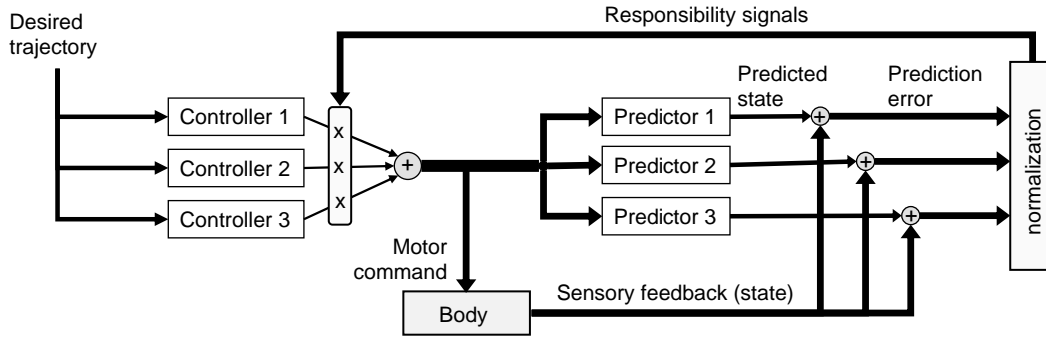
Fig. 11. The functioning of the MOSAIC model in the control mode. The responsibility signals indicate how well the control modules are suited for the control task at hand. The overall control output is the sum of the output of controller modules as weighted by the responsibility signals.

The aim of motor control is to produce motor commands $\Gamma(t)$ at time t such that a desired state[9] $x_{\text{des}}(t)$ is attained by the controlled system dynamics ψ. The net motor output Γ of the MOSAIC model is determined by a set of adaptive controller–predictor pairs (ψ_i,ϕ_i) via the responsibility signals λ_i which are computed using the predictor outputs and the current state of the system. The equations given in Table 1 describe the control mechanism more rigorously (for simplicity, we use a discrete time representation). The adaptive nature of the controller–predictor pairs is shown with semicolons as $\psi_i(\cdot;w_i)$ $\phi_i(\cdot;\upsilon_i)$ meaning that ψ_i and ϕ_i are functions determined by the parameters w_i and υ_i, which are typically the weights of a function approximator or a neural network.

Rather than presenting the details of how the controller–predictor pairs can be adapted (trained) for a variety of the tasks we note that MOSAIC is described without strict attachment to a particular learning method, so it is possible to derive various learning algorithms for adapting controller–predictor pairs. In particular, gradient descent (Wolpert & Kawato, 1998) and expectation maximization (Haruno et al., 2001) learning algorithms have been derived and applied for motor control learning.

12.1. Imitation and action recognition with MOSAIC

Although MOSAIC was initially proposed for motor control, it is possible to utilize it for imitation and action recognition. This dual use of the model establishes some parallels between the model and the mirror neuron system. The realization of imitation (and action recognition) with MOSAIC requires three stages: first, the visual information of the actor's movement must be converted into a format that can be used as inputs to the motor system of the imitator (Wolpert et al., 2003). This requires that the visual processing system extracts variables akin to state (e.g. joint angles) which can be fed to the imitator's MOSAIC as the 'desired state' of the demonstrator (Wolpert et al., 2003). The second stage is that each controller

generates the motor command required to achieve the observed trajectory (i.e. the desired trajectory obtained from the observation). In this 'observation mode', the outputs of the controllers are not used for actual movement generation, but serve as input to the predictors paired with the controllers (see Fig. 12). Thus, the next likely states (of the observer) become available as the output of the forward predictions. These predictions then can be compared with the demonstrator's actual next state to provide prediction errors that indicate, via responsibility signals, which of the controller modules of the imitator must be active to generate the movement observed (Wolpert et al., 2003).

12.2. Relation to MNs and imitation

The responsibility signals are computed by (softmax) normalizing the prediction errors as shown in Fig. 12. Notice that the responsibility signals can be treated as symbolic representation describing the observed (continuous) action. The temporal stream of responsibility signals representing the observed action then can be used immediately or stored for later reproduction of the observed action. Simulations with the MOSAIC model indicate that the aforementioned imitation mechanism can be used to imitate the task of swinging up a one degree of freedom jointed stick against gravity through observation of successful swing ups (Doya et al., 2000).

Within the MOSAIC framework the output of predictors might be considered analogous to mirror neuron activity. This would be compatible with the view of the MSI model, where it is suggested that the mirror neurons may implement a motor-to-sensory forward model. It is however necessary to point out one difference. The MSI model deals with motor control relying only on visual input and kinematics (leaving the

___
[9] The term 'state' generally represents the vector of variables that are necessary to encapsulate the history of the system as a basis for describing the system's response to the external inputs, which then involves specification of current output and of the updating of the state. For a point mass physical system the state combines the position and velocity of the mass.

Table 1
The equations describing the control function of the MOSAIC model

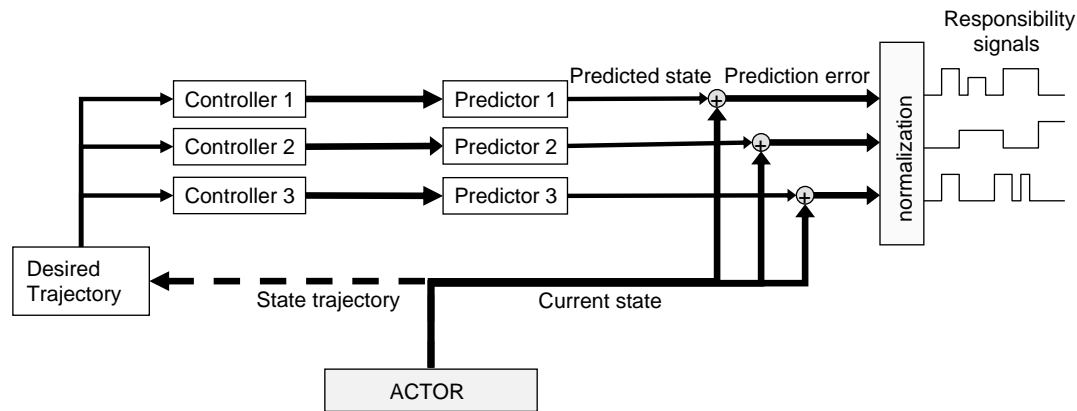|	
Dynamics of the controlled system	$x(t+1) = \Psi(x(t),\Gamma(t))$
The MOSAIC (net) control output	$\Gamma(t+1) = \sum_i \lambda_i\psi_i(x_d(t+1),x(t);w_i)$
Individual controller (inverse model) outputs	$u_i(t) = \psi_i(x_d(t+1),x(t);w_i)$
Individual predictions (forward model)	$\hat{x}_i(t+1) = \phi_i(x(t),u(t);\upsilon_i)$
The responsibility signals	$\lambda_i(t) = \dfrac{e^{(x(t)-\hat{x}_i(t))^2/\delta^2}}{\sum_k e^{(x(t)-\hat{x}_k(t))^2/\delta^2}}$

Fig. 12. The functioning of the MOSAIC model in the observation mode is illustrated. For imitation, the responsibility signals indicate which of the controller modules must be active to generate the movement observed.

dynamics to the lower motor areas). In contrast, MOSAIC is a true control system that deals with dynamics. Thus, the nature of forward models in the two architectures is slightly different. The output of the forward model required by MSI is in visual-like coordinates (e.g. the orientation difference of the hand axis and the target object), whereas for MOSAIC the output of the forward models are more closely related to the intrinsic variables of the controlled limb (e.g. joint position and velocities). However, it is possible to envision an additional dynamics-to-visual forward model that can take MOSAIC forward model output and convert it to some extrinsic coordinates (e.g. distance to the goal). In a way, the MSI's forward model can be such an integrated prediction circuit implemented by several brain areas.

A final note here is that the internal models envisioned in the neuroscience literature (e.g. Carr et al., 2003; Iacoboni et al., 1999; Miall, 2003) are usually at a much higher level than the internal models of the MOSAIC or the MSI model introduced here, which are much harder to learn from a computational point of view.

13. A Taxonomy of models based on modeling methodology

When the system to be modeled is complex it is often necessary to focus on one or two features of the system in any one model. The focus of course is partly determined by the modeling methodology followed by the modeler. Here, we present a taxonomy of modeling methodologies one can follow, and compare the models we have presented accordingly.

The utility of a model increases with its generality and ability to explain and predict observed and unobserved behavior of the system modeled. The validity and utility of a model is leveraged when all the known facts are incorporated into the model. This is called *data driven* modeling where the modeler's task is to develop a computational mechanism (equations and computer simulations) that replicates the observed data with the hope that some interesting, non-trivial predictions can be made. This is the main modeling approach for cellular level neuron modeling. Although one would expect

that the neurophysiological data collected so far is to be widely used as the basis for computational modeling, we unfortunately lack sufficient quantitative data on the neurophysiology of the mirror system. Most mirror system (related) modeling *assumes* the *generic* properties of mirror neurons to build imitation systems rather than addressing hard data. We have included this type of model in the taxonomy because there is certain utility of those models as they lead to questions about the relation between mirror neurons and imitation.

Models based on evolutionary algorithms have been used in modeling the behaviors of organisms, and developing neural circuits to achieve a prespecified goal (e.g. central pattern generators) in a simplified simulated environment. Although this type of modeling, in general, does not make use of the available data, the model of Borenstein and Ruppin (2005) suggests how mirror neurons might have come to be involved in imitation or other cognitive tasks. However, as we have already noted, 'real' evolution may have exapted the mirror system to support imitation, rather than starting from the need to imitate and 'discovering' mirror neurons as a necessary tool. An evolutionary point of view can also be adopted to build models that do not employ evolutionary algorithms, which start off by postulating a logical *reason for the existence* of the mirror neurons. The logic can be based on the location of the mirror neurons, or on the known general properties of neural function. The former logic dictates that mirror neurons must be involved in *motor control*. The latter logic (*phenomenal*) draws on Hebbian plasticity mechanisms and dictates that represen-tations of contingent events are associated in the cerebral cortex. Fig. 13 illustrates how the models we have presented fall into our taxonomy. However, note that this taxonomy should not be taken as defining sharp borders between models. Models focusing on imitation can be cast as being developed following a 'reason of existence'. For example DRAMA architecture (Billard & Hayes, 1999) employs a Hebbian like learning mechanism and thus can be considered in the ASSOC category in Fig. 13. Similarly, although no motor control role was emphasized for mirror neurons in Demiris's imitation system (Demiris & Johnson, 2003), it employs mechanisms similar to those of the MOSAIC model.
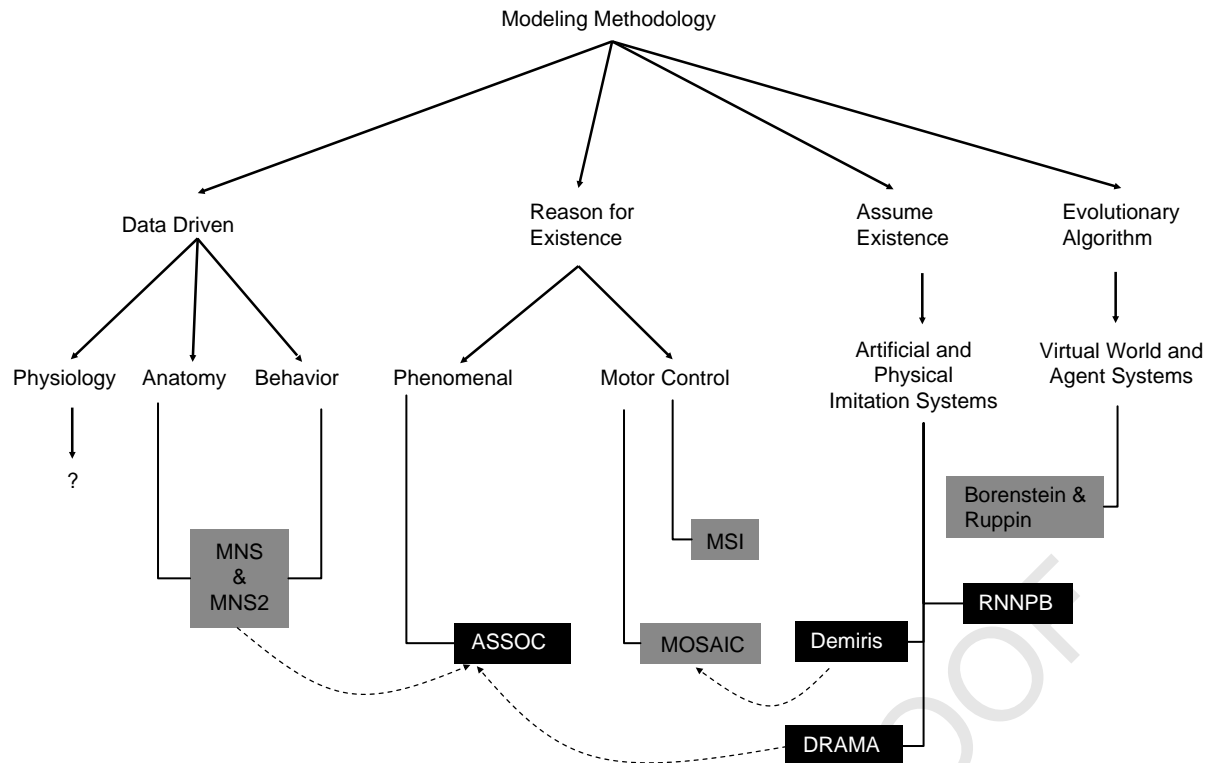
Fig. 13. A taxonomy of modeling methodologies and the relation of the models presented. Dashed arrows indicate that the linked models are similar or can be cast to be similar (see text).

Table 2
A very brief summary of models presented in terms of biological relevance, architecture and the relevant results and predictions

	Biological relevance	Architecture	Results/predictions
RNNBP	N/A (general)	Distributed representation Recurrent network with a complex generative model	Imitation, action recognition There should be MN inhibition for undesired action imitation
MOSAIC and Demiris	N/A (general)	Modular, localist Demiris: no module learning MOSAIC: Modules learn with EM or gradient based learning	Modular control allows imitation and action recognition MN response limited to certain speed range Existence of subaction encoding units
Borenstein and Ruppin	N/A (evolutionary model)	Virtual world, agent system	Imitation and MN formation is favored by natural selection
MSI	Based on general organization of system level anatomy (premotor and parietal)	Connectionist schemas	Mental state Inference Kinematics of observed action should correlate with mirror activity
ASSOC and DRAMA	N/A (general. However, Billard & Mataric, 2003 embedded DRAMA in a biologically inspired imitation architecture)	Recurrent, Hopfield like	Imitation and MN formation Self-observation may be involved in formation of MNs and development of imitation
MNS	Uses system level anatomy (premotor and parietal) and addresses mirror neuron firing	Hidden layer back-propagation network with supporting schemas	Prediction on neural firing patterns of MNs for: Altered kinematics Spatial mismatch Novel object grasping
MNS2	Uses system level anatomy (premotor and parietal) and addresses mirror neuron firing	Recurrent back-propagation network for visual input; Hebbian synapses for auditory input; working memory and dynamic remapping	Addresses data on mirror neurons for grasping including audio–visual mirror neurons and response to grasps with hidden end-state

We now contrast and define common characteristic of some of the models presented in this article with the hope that future (especially data driven) modeling can embrace a larger set of mirror neuron functions. Note that the taxonomy we present is orthogonal to the granularity of the modeling (i.e. cell, cell population, brain region or functional/abstract). The MOSAIC model in its original form does not advocate any brain structure for its functional components. Similarly the associative memory models of the mirror system (ASSOC in Fig. 13) in general do not strongly adhere to any brain area. Therefore, we can consider them as 'general models'. On the other hand, MSI and—to a larger extent—MNS associate (macaque) brain regions with the functional components. The two models are compatible in terms of brain areas supporting the mirror system. MNS conceptually accepts that mirror neurons must have a role in motor control. However, it does not provide mechanisms to simulate this because the focus of the modeling in MNS is the development of mirror neurons. (MNS 2 extends MNS by addressing data on audiovisual mirror neurons and on grasps with hidden end-state). The key to both MNS and MSI is the *object-centered representation* of the actions and the *self-observation* principle, which together allow actions to be recognized irrespective of the agent performing it. However, there is one difference: in MSI, self-observation is purposeful; it is used to implement a visual feedback control loop for action execution. On the other hand, the MNS model, despite the differences in the adaptation mechanisms, resembles the associative memory models that operate on the principle of association of stimuli produced as the result of movement execution. The purposeful self-observation (visual feedback) principle of the MSI model establishes the logical link between the MSI and the MOSAIC model in spite the differences of two models in terms of motor control ability. The MOSAIC model is crafted for true motor control (i.e. it considers dynamics) whereas MSI deals only with kinematics. Table 2 provides a succinct account of the presented models by enlisting the key properties and results relevant for this article.

14. Conclusion

In spite the accumulating evidence that humans are endowed with a mirror system (Buccino et al., 2001; Hari et al., 1998; Iacoboni et al., 1999), it is still an open question how our brains make use of this system. Is it really used for imitation or mental state estimation? Or is it simply an action recognition system? A predominant assumption among computational modelers is that human mirror neurons subserve imitation. Although there are many imitation models based on this view, there are virtually no studies addressing the assumption itself. We need biologically grounded computational models to justify this view; the models of this sort should address the computational requirements and the possible evolutionary changes in neural circuitry necessary to allow mirror neurons to undertake an important role for imitation.

Our view is that, the location of mirror neurons (MNs) indicates that the function of mirror neurons must be rooted in motor control. We emphasize that future computational models of MNs that share the same view, regardless of whether they address imitation or not, must explain the dual role of the mirror system by showing computationally that MNs perform a useful function for motor control. Note that this need not mean that all mirror neurons in the human brain that are 'evolutionary cousins' of macaque mirror neurons for grasping need themselves be involved in manual control—one can accept that mirror neurons for language have this cousinage without denying that lesions can differentially yield aphasia and apraxia (see Barrett, Foundas, & Heilman, 2005, and also the Response in Arbib, 2005).

Imitation is just one way to look at the problem; the mirror neurons must also be analyzed and modeled in an *imitation-decoupled* way. For example, is it possible that MNs are simply the consequence of Hebbian learning, i.e. an automatic association of the corollary discharge and the subsequent sensory stimuli generated as the associative memory hypothesis claims? The MNS (Oztop & Arbib, 2002) and MNS2 (Bonaiuto et al., 2005) models postulate that some extra structure is required, both to constrain the variables relevant for the system, and to track trajectories of these relevant variables. Most models assume that the relevant variables are indeed supplied as input. The work of Kumashiro et al. (2003) reminds us that in fact the mirror system must be augmented by an attentional system that can ensure that the appropriate variables concerning agent and object are made available.

The MNS model shows how mirror neurons may learn to recognize the hand-state trajectories (hand-object relationships) for an action already within the repertoire. We have made clear that extra machinery is required to go from a novel observed action to a motor control regime which will satisfy it. MOSAIC approaches this by generating a sequence of responsibility signals representing the observed action as a string of segments of known actions whose visual appearance will match the observed behavior. Arbib and Rizzolatti (1997) set forth the equation Action = Movement + Goal to stress the importance of seeing movements in relation to goals, rather than treating them in isolation. Arbib (2002) then argued that humans master *complex imitation* by *approximating* novel actions by a combination (sequential as well as co-temporal) of known actions, and then making improvements both by attending to missing subactions, and by tuning and coordinating the resulting substructures. This is much the same as what Wohlschlager, Gattis, and Bekkering (2003) call *goal-directed imitation*. On their view, the imitator does not imitate the observed movement as a whole, but rather decomposes it into hierarchically ordered aspects (Byrne & Russon, 1998), with the highest aspect becoming the imitator's main goal while others become subgoals. The main goal activates the motor schema that is most strongly associated with the achievement of that goal. Of course, there is no 'magic' in complex imitation which automatically yields the right hierarchical decomposition of a movement. Rather, it may be the success or failure of a 'high-level approximation' of the observed action that leads to attention to crucial subgoals which were not observed at first, and thus leads, perhaps somewhat circuitously, to successful

imitation. But the point remains that this process is in general much faster than the time-consuming extraction of statistical regularities in Byrne's (2003) 'imitation by (implicit) behavior parsing'—we might refer to complex (goal-directed) imitation as imitation by *explicit* behavior parsing. Finally, we note that this process is postulated to result in a new motor schema (forward + inverse model) which may be linked to those previously available, but which now constitutes a new action which is henceforth available to further refinement of inverse and forward models separate from those which have been acquired before.

Although a decade has passed since the first reports of mirror neurons came out, the reciprocal lack of full knowledge (or interest) between the sides of conceptual and computational modeling is evident. To close this gap experimentalists should conduct experiments that involve quantitative measurement, e.g. relating neuronal activity to synchronized kinematics recordings of the experimenter and monkey during action demonstration and execution. This could shed light on the debate between those who believe a mirror neuron encodes a specific action, and those who seek to understand how mirror neurons may provide population codes for action-related variables. The correlation between the discharge profiles of mirror neurons with various visual feedback parameters will provide modelers invaluable information to construct models that capture neurophysiological facts. Likewise, we do not know anything about the developmental stages of mirror neurons. Are they innate? Probably not, so what circuitry and adaptation mechanisms are involved? To get answers for these questions, computational modeling that can provide a causally complete account of mirror neurons and the larger system of which they are part is crucial. Only then would it be possible to develop models of motor control, imitation, mental state inference, etc. that assign various roles to mirror neurons and their interactions with diverse brain regions in an empirically justified way.

Acknowledgements

References

Arbib, M. (2002). The mirror system, imitation, and the evolution of language. In C. Nehaniv, & K. Dautenhahn (Eds.), *Imitation in animals and artifacts* (pp. 229–280). Cambridge, MA: MIT Press.

Arbib, M., & Rizzolatti, G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication and Cognition*, *29*, 393–424.

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. B. Brooks (Vol. Ed.), *Handbook of physiology, section 2: The nervous system. Motor control, part 1: Vol. II* (pp. 1449–1480). Bethesda, MD: American Physiological Society.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(2), 105–124 (discussion 125–167).

Arbib, M. A., Erdi, P., & Szentagothai, J. (1998). *Neural organization: Structure, function, and dynamics*. Cambridge, MA: MIT Press.

Barrett, A. M., Foundas, A. L., & Heilman, K. M. (2005). Speech and gesture are mediated by independent systems. *Behavioral and Brain Sciences*, *28*, 125–126.

Billard, A., & Hayes, G. (1999). DRAMA, a connectionist architecture for control and learning in autonomous robots. *Adaptive Behavior*, *7*(1), 35–63.

Billard, A., & Mataric, M. J. (2001). Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture. *Robotics and Autonomous Systems*, *37*(2–3), 145–160.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567.

Bonaiuto, J., Rosta, E., & Arbib, M. A. (2005). Recognizing invisible actions, workshop on modeling natural action selection. Workshop on modeling natural action selection, Edinburgh.

Borenstein, E., & Ruppin, E. (2005). The evolution of imitation and mirror neurons in adaptive agents. *Cognitive Systems Research*, *6*(3).

Brass, M., Bekkering, H., Wohlschlager, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, *44*, 124–143.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*(2), 400–404.

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, *89*(2), 370–376.

Byrne, R. W. (2003). Imitation as behaviour parsing. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences*, *358*(1431), 529–536.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, *21*(5), 667–684 (discussion 684–721).

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *PNAS*, *100*(9), 5497–5502.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904.

Demiris, Y., & Hayes, G. (2002). Imitation as a dual-route process featuring predictive and learning components: A biologically-plausible computational model. In K. Dautenhahn, & C. Nehaniv (Eds.), *Imitation in animals and artifacts*. Cambridge, MA: MIT Press.

Demiris, Y., & Johnson, M. (2003). Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science*, *15*(4).

Doya, K., Katagiri, K., Wolpert, D., & Kawato, M. (2000). Recognition and imitation of movement patterns by a multiple predictor–controller architecture. Technical Report of IEICE TL2000-11, (33–40).

Elshaw, M., Weber, C., Zochios, A., & Wermter, S. (2004). An associator network approach to robot learning by imitation through vision, motor control and language. International joint conference on neural networks, Budapest, Hungary.

Fadiga, L., & Craighero, L. (2004). Electrophysiology of action representation. *Journal of Clinical Neurophysiology*, *21*(3), 157–169.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*(2), 399–402.

Fogassi, L., Gallese, V., Buccino, G., Craighero, L., Fadiga, L., & Rizzolatti, G. (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey—a reversible inactivation study. *Brain*, *124*, 571–586.

Fogassi, L., Gallese, V., Dipellegrino, G., Fadiga, L., Gentilucci, M., Luppino, G., et al. (1992). Space coding by premotor cortex. *Experimental Brain Research*, *89*(3), 686–690.

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

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15061–15065.

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, *13*(10), 2201–2220.

Hassoun, M. (1993). *Associative neural memories: Theory and implementation*. New York: Oxford University Press.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), e79.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526–2528.

Ijspeert, A., Nakanishi, J., & Schaal, S. (2003). Learning attractor landscapes for learning motor primitives. In S. Becker, S. Thrun, & K. Obermayer (Vol. Eds.), *Advances in neural information processing systems: Vol. 15* (pp. 1547–1554). Cambridge, MA: MIT Press.

Ito, M., & Tani, J. (2004). On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system. *Adaptive Behavior*, *12*(2), 93–115.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, *3*, 79–87.

Kawato, M., & Gomi, H. (1992). A computational model of 4 regions of the cerebellum based on feedback-error learning. *Biological Cybernetics*, *68*(2), 95–103.

Kawato, M., & Wolpert, D. (1998). Internal models for motor control. *Sensory Guidance of Movement*, *218*, 291–307.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*, 522–525.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*(5582), 846–848.

Kumashiro, M., Ishibashi, H., Uchiyama, Y., Itakura, S., Murata, A., & Iriki, A. (2003). Natural imitation induced by joint attention in Japanese monkeys. *International Journal of Psychophysiology*, *50*(1–2), 81–99.

Kuniyoshi, Y., Yorozu, Y., Inaba, M., & Inoue, H. (2003). From visuo-motor self learning to early imitation—a neural architecture for humanoid learning. International conference on robotics & automation, Taipei, Taiwan.

Miall, R. C. (2003). Connecting mirror neurons and forward models. *Neuroreport*, *14*(17), 2135–2137.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279.

Morita, M. (1996). Memory and learning of sequential patterns by nonmonotone neural networks. *Neural Networks*, *9*(8), 1477–1489.

Murata, A. (2005). Function of mirror neurons originated from motor control system (in Japanese). *Journal of Japanese Neural Network Society*, *12*(1), 52–60.

Myowa-Yamakoshi, M., & Matsuzawa, T. (1999). Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *113*(2), 128–136.

Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, *87*(2), 116–140.

Oztop, E., Chaminade, T., Cheng, G., Kawato, M. (2005). Imitation bootstrapping: Experiments on a robotic hand. IEEE-RAS international conference on humanoid robots, Tsukuba, Japan.

Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Brain Research Cognitive Brain Research*, *22*(2), 129–151.

Rizzolatti, G. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*(3), 491–507.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology (Berl)*, (1–3).

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670.

Schaal, S., Ijspeert, A., & Billard, A. (2003). Computational approaches to motor learning by imitation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 537–547.

Schaal, S., Peters, J., Nakanishi, J., & Ijspeert, A. (2004). Learning movement primitives. International symposium on robotics research, Ciena, Italy.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage*, *25*(1), 76–89.

Tani, J., Ito, M., & Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot experiments using RNNPB. *Neural Networks*, *17*(8–9), 1273–1289.

Tunik, E., Frey, S. H., & Grafton, S. T. (2005). Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp. *Nature Neuroscience*, *8*(4), 505–511.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al. (2001). I know what you are doing: A neurophysiological study. *Neuron*, *31*(1), 155–165.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664.

Wohlschlager, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: An instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 501–515.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 593–602.

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, *5*(11), 487–494.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.