Faculty of Mathematics, Physics, and Informatics
Comenius University, Bratislava

Mirror neurons – theoretical and computational issues

Igor Farkaš, Michal Malý, Kristína Rebrová

TR-2011-028

Technical Reports in Informatics

Mirror neurons – theoretical and computational issues

Igor Farkaš, Michal Malý, Kristína Rebrová
Department of Applied Informatics
Comenius University in Bratislava

Abstract

The discovery of mirror neurons, first in macaques and most recently in humans, has provided an interesting research agenda on the possible roles of mirror neurons in various cognitive functions. Compared to that of monkeys, the functionality of the human mirror system seems more general and more abstract. However, the most recent views on mirror neurons point to disagreements regarding their development and functions in human cognition. In this review, we start with some of the theoretical points of disagreement and propose their reconciliation. Namely, we argue for a position between the "adaptation" hypothesis (emphasis on inborn biases) and the "association" hypothesis (emphasis on learning) of the origin of mirror neurons. Then we propose a graded action understanding hypothesis, to be placed between the motor theory and the perceptual theory of action understanding. We explain justification of our hypothesis in the extended context of actions whose "deep understanding" requires special training. In addition, we argue that the role of STS, subserving action recognition, has not been fully appreciated in the mirror neuron theory. In the second part, we critically review conceptual and computational models of the mirror neuron system that range from the simplest, Hebbian accounts, to more complex accounts, according to the interaction between sensory and motor representations. We propose that a plausible computational model of action understanding should account for acquisition of a mirroring property and include mechanisms for yielding high-level view-invariant perceptual representations needed for a neural account of action recognition.

Keywords: mirror neurons; action understanding; sensory/motor representations; computational modeling

1 Introduction

The discovery of mirror neurons in macaques (Pellegrino et al., 1992) has sparked a lot of scientific interest that led to the formulation of various theoretical positions, conceptual models and computational models. The very recent evidence supports the existence of mirror neurons also in humans (Mukamel et al., 2010). The prevailing view of the human mirror system (MNS), hypothesizing its role in a variety of cognitive functions including action understanding, imitation, empathy and mindreading, has recently been questioned by some researchers, claiming that mirror neurons are only a byproduct of associative learning (e.g. Heyes, 2010; Hickok, 2008). What is therefore the current state-of-the-art regarding the mirror neuron theory?

In finding neural substrates of action understanding, two major dichotomies can be identified. First, there is a dichotomy between the actual mechanism mediating understanding of perceived actions (Rizzolatti et al., 2001). The *visual hypothesis*, according to which understanding is based solely on visual assessment of various elements of the scene, and *direct-matching* *hypothesis*, based on the discovery of mirror neurons, which states that resonance in motor areas of the brain is necessary for the understanding of the observed action. The second dichotomy regards the origin of mirror neurons as well as their role in the emergence of high-level cognitive functions such as language. On one hand, mirror neuron theory posits, that mirror neurons are an evolutionary adaptation and that their function is somehow predetermined (according to one of its critics (Heyes, 2010)). Rizzolatti and Arbib (1998) propose, that mirror neurons had played a crucial role in the development of high-level social skills such as language and therefore were evolutionarily imporant. On the other hand, the association hypothesis (Heyes, 2010) claims that mirror neurons are only a by-product of highly correlated inputs that emerge during self-observation, and that they play a little, if any role in cognition. Lastly, it is important to note that the action understanding ability appears to be comprehended differently by different researchers and sometimes it is equated with action recognition.

The aim of this paper is to address the two main dichotomies and propose a reconciliation for both of them, based on a new view on action understanding. We propose the *graded action understanding hypothesis*, claiming that both mere action recognition and action understanding are two degrees of one phenomenon, with different demands on various processing streams of the brain. Unlike Rizzolatti and Sinigaglia (2010), who answer to the criticism of the mirror neuron theory based on dissociation studies (Mahon and Caramazza, 2005) with a proposal of a dual-processing theory, our hypothesis covers both visual and direct-matching accounts unitedly assigning them different roles in the process of action understanding. However, we maintain the general assumption of mirror neuron theory that deep understanding, unlike shallow understanding or mere recognition, necessarily involves the activation in motor areas of the brain.

The paper is structured as follows. First we review the empirical evidence regarding mirror neurons (section 2). In section 3 we focus on their possible roles in human cognition. In section 4, we propose a reconciliation of the adaptation and association hypotheses of the origin of mirror neurons. Then we focus on action understanding capacity (section 5) and provide our view, which is based on recent empirical evidence. Next we switch to computational issues and in section 6 we provide a critical assessment of various computational and conceptual models that deal with mirror neurons. In a discussion (section 7), we review theoretical and computational aspects that we find important and/or interesting. Section 8 briefly concludes the paper.

2 Empirical evidence

2.1 Mirror neurons in monkeys

Mirror neurons were originally discovered in ventral premotor cortex of the macaque monkey – the area F5. This area is characteristic with neurons that become active during particular hand movements (such as grasping, holding and tearing) and mouth movements. Many of these neurons react only to very specific types of actions (e.g. only to a precision grip) and some of them are activated by visual stimuli (Rizzolatti et al., 1988). Pellegrino et al. (1992) reported that some of these neurons discharged not only during the execution of a certain motor act, but also when the monkey observed the particular motor act performed by the experimenter, provided that the target of the motor act was present on the scene and the motor act finished completely (a piece of food was grasped by the experimenter). Mirror neurons did not respond to meaningless actions or to the presentation of an object alone (even to food). These specifically oriented neurons and their properties were thoroughly described by Gallese et al. (1996) and

Rizzolatti et al. (1996).

According to the first findings, mirror neurons were considered to be involved specifically in action recognition. This function was attributed to so called *parieto-frontal action-observation action-execution brain circuit*[1], which consists of the area F5 (premotor cortex), area PFG, located in rostral part of inferior parietal lobule (IPL) between areas PF and PG, and the anterior intraparietal area (AIP). The latter two areas are both connected with F5, but also receive high-order visual information from areas located inside the superior temporal sulcus (STS) and the inferior temporal lobe (IT). The STS, similarly to F5, encodes biological motion, but it lacks motor properties and therefore cannot be considered a true part of the mirror neuron system (along with the IT). The parieto-frontal circuit is also connected with the area F6 (pre-supplementary motor area) and the ventral prefrontal cortex, which are the higher-order areas that control it.

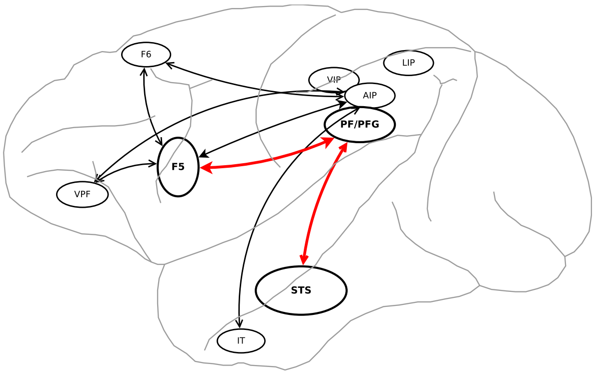

Figure 1: A schematic depiction of relevant areas of macaque brain.

In addition to the parieto-frontal circuit, neurons with mirror properties were found in other areas of the parietal lobe, the lateral intraparietal area (LIP), which contributes to joint attention (Shepherd et al., 2009), and ventral intraparietal area (VIP). Neurons in VIP encode tactile and visual stimuli occurring in peripersonal space and might be responsible for encoding body-directed motor acts rather than object-directed motor acts represented by mirror neurons in F5 (Ishida et al., 2010). Mirror neurons were discovered also in frontal areas of a monkey brain, in primary motor (M1) and dorsal premotor (PMd) cortices (Tkach et al., 2007). The fact that neurons with mirror properties were discovered in various parts of the monkey brain does not, we think, diminish their importance, as argued by some authors (Heyes, 2010). On the contrary, it is likely that they mediate perceptual simulation (Barsalou, 1999) which facilitates understanding in various modalities. Although the action-execution action-observation circuit was discovered first and remained in the center of attention, it is likely that the function of mirror neurons is not limited to action understanding/recognition.

2.2 Mirror neurons in humans

Although the direct evidence for mirror neurons in humans emerged only recently, the background for assuming their existence is more than 50 years old. In the 1950s, Gastaut and

[1] This name has been suggested only lately by Rizzolatti and Sinigaglia (2010).

colleagues (Cohen-Seat et al., 1954; Gastaut and Bert, 1954) observed that the desynchronization of the EEG mu rhythm, characteristic for the motor rest, occurs not only during active movements of studied subjects, but also when the subjects observed actions performed by others, or even a robotic arm (Oberman and Ramachandran, 2007). This evidence was confirmed by various EEG, MEG, and TMS studies, for a detailed summary see Rizzolatti and Craighero (2004). According to this study, the core of the human MNS consists of the rostral part of the IPL, the lower part of the precentral gyrus, and the posterior part of the inferior frontal gyrus (IFG).

The first proof of mirror neurons in humans on a single-cell level was provided by Mukamel et al. (2010). The study contains recordings from 21 patients with pharmacologically intractable epilepsy, who executed or observed hand grasping actions and facial emotional expressions. The observed regions of the brain were chosen according to clinical criteria only, so the main areas of interest, namely the Broca's area, were not examined. Significant proportions of cells responding to both action observation and action execution were found in the medial frontal lobe (SMA) and, interestingly, in the medial temporal lobe (namely in the hippocampus, the parahippocampal gyrus and the entorhinal cortex). According to Mukamel et al. (2010), the mirroring activity, recorded during observation of an action in the medial temporal lobe, might correspond to the reactivation of the memory traces formed previously during its execution. This study clearly shows that the function of mirror neurons varies according to their location in the brain.

Another striking finding here was the discovery of a subset of mirror neurons with opposite patterns of excitation and inhibition during observation versus execution of an action. This mechanism might be responsible for maintaining the sense of self/others differentiation, and for the control of unwanted imitation during action observation. The similar mechanism was discovered in monkeys (Kraskov et al., 2009).

Unlike the mirror neurons in monkeys, which are triggered only by meaningful actions like reaching and grasping of food, human mirror neurons react also to intransitive meaningless arm movements (Fadiga et al., 1995). Rizzolatti and Sinigaglia (2010) conclude that additionally to motor acts, the human mirror system also encodes sole body movements from which the motor acts and actions are built, what enables a sort of "parsing mechanism" for a better understanding of complex motor acts. Another crucial difference between humans and monkeys is, that human mirror system, unlike that of the monkey, responds to actions regardless of the effector used to perform them, be it an animal, a human, or a robotic arm, or even a tool, with or without the presence of the target object (Peeters et al., 2009). The mirror system in monkey, on the other hand, requires an interaction between a biological effector (hand or mouth) and an object, and will not respond to an agent mimicking an action (Rizzolatti and Craighero, 2004). This suggests that the human mirror system is more general and more abstract, lending itself to supporting a larger variety of cognitive functions.

3 The role of mirror neurons in cognition

According to Rizzolatti and Arbib (1998), but also other researchers, the activity of mirror neurons represents actions for two purposes – imitation and understanding of actions. We will describe the neural basis and main theories on the role of mirror neurons in these two important cognitive abilities and also briefly review some other functions described in the literature, namely goal-encoding, empathy and language.

3.1 Mirror neurons and action understanding

There have been several definitions of action understanding formulated, here we follow Gallese et al. (1996), according to whom it is a capacity to recognize that another individual is performing an action, to differentiate it from other actions and to act appropriately on the basis of this information. Two major hypotheses of action understanding have been proposed to date, the classical *visual hypothesis* and the *direct-matching hypothesis* (Rizzolatti et al., 2001) that represent largely contrasting views regarding the localization of function and the information flow. According to the visual hypothesis, action understanding is based solely on visual analysis of different elements of the scene. On the contrary, the direct-matching hypothesis assigns a more central role to the motor information in perceptual (mostly visual) tasks.

Although the direct-matching hypothesis is plausible in many situations, there are still examples, as its critics argue, in which it does not hold, for instance when an individual observes another playing a music instrument. Hickok (2008) points out that the observer does not need any motor skill necessary to play saxophone, in order to know that the observed individual is playing the saxophone. As a reaction to this argument, Rizzolatti and Sinigaglia (2010) propose a kind of dual-processing theory describing two types of action understanding: motor and non-motor-based understanding, the latter of which implies merely semantic knowledge about the world and provides the observer with superficial information about the observed motor act.

This distinction is closely related to the opposing views (disembodied and embodied) on the organization of conceptual representations in the brains, depending on the role of sensorimotor systems. Embodied theories claim that conceptual content is reducible to sensorimotor content, whereas disembodied theories do not.

3.2 Mirror neurons and imitation

Heyes (2001), refers to imitation as to “copying by an observer of a feature of the body movement of a model” implying a causal link between observation of the movement and its execution. The imitation of the particular movement feature must happen after observation of this (and not other) feature, and it must not occur by chance. A similar position is taken by Rizzolatti et al. (2001) who emphasize the “response facilitation” – an automatic tendency to reproduce the observed movement, which may occur without understanding or with understanding, the latter being present only in adult humans and mediated by mirror neurons.

A more detailed description and, more importantly, different distinctions were introduced by Hurley (2008). She distinguishes four types of social learning: stimulus enhancement, goal emulation[2], movement priming and true imitation, while the latter is only present in humans, and as shown recently, also in some “higher” animals like birds and monkeys; for a review see (Iacoboni, 2009). True imitation in this sense does not only require a proper copying of movements (means), but also a successful copying of goals (ends). Hurley explains how the ability to imitate, but also deliberation and mindereading, can be enabled by subpersonal (functional, not conscious, nor neural) mechanisms of control, mirroring, and simulation in her multi-layer Shared Circuits Model (SCM). In this model, consistently with the embodied cognition school,

[2] The term “goal emulation” refers to a situation when an animal (or a human) sees an action and then produces a slightly different action to produce the same goal. It is known that children imitate both means and goals “blindly”, even when the means are ineffective. On the other hand, it was found that children, given a context, will copy the goal, but using a simpler movement to achieve it, than the one executed by the experimenter (Bekkering et al., 2000).

the ideomotor theory of action (James, 1890), and consequently with the direct-matching hypothesis, perception and action share the same neural resources. Hurley's aim is to reconcile ideomotor and associative theories of action understanding that are often posed against each other, emphasizing both the sharing of resources and associative learning.

An interesting question arises considering the development of imitative abilities in humans. Both infants and their parents are known to imitate facial gestures due to overt (contagious) imitation processes. As proposed by Heyes (2010), the imitative behavior of parents might as well serve as a natural mirror for infants to associate their gestures with their visual representations, providing then the motor resonance underlying action understanding. This can be considered a "imitation-first" view. On the other hand, "understanding-first" view emphasizes that understanding is needed in order to imitate. In Hurley's view, which we find plausible, these two skills develop hand-in-hand. At first, even if the imitation is not successful, for instance when the movement is somehow copied but its goal is not achieved, neural representation of the new movement starts to form. Following multiple attempts, this representation gains its strength as well as the motor resonance evoked during the re-observation needed for further attempts to imitate, since a similar motor plan is already in the repertoire. The motor resonance can also explain overt imitation, like contagious yawning. It can be viewed as an uncontrolled mirror response. On the other hand, better understanding of the observed action, mediated by mirror neuron response elicited by similarity of parts of the observed action with motor plans already present in observer's motor repertoire, might also facilitate the process of imitation. In this sense, imitation and action understanding (mediated by mirroring processes) develop together in a mutually beneficial way.

3.3 Mirror neurons and goal understanding

Another important point made by Rizzolatti and Sinigaglia (2010) is that "mirror neurons encode the goal of the observed motor act", regardless of the effector or circumstances in which it is observed, so there is no need for strict matching of the observed act with some other act from the observer's motor repertoire. The intuition behind this theory is the existence of two types of mirror neurons according to their congruence with the observed action. The *strictly congruent* mirror neurons react only to certain type of motor act, for instance only to a precision grip. The *broadly congruent* mirror neurons, on the other hand, may react to a whole category of motor acts sharing the same goal.

Compelling evidence for the goal-encoding role in monkeys was provided by Umiltà et al. (2008) who recorded single-cell activity in monkeys using and observing the usage of two types of pliers (normal and reverse pliers with opposite mechanisms for opening and closing). Results of this study showed the same pattern of mirror neurons' activity during the observation of both types of grasp. According to Rizzolatti and Sinigaglia (2010), the evidence for goal-encoding in humans lies in the broad congruency of the effector used to perform the motor act (described in section 3.1).

Another example supporting the goal-encoding hypothesis was provided in an interesting fMRI study (Gazzola et al., 2007) with aplasic individuals (born without arms) who observed actions performed by hands, feet, and mouth. The results of this experiment showed the mirroring activity also for hand actions (that subjects were never able to execute), goals of which they were able to accomplish by mouth or feet. These interesting findings suggest that there exist broadly congruent goal-encoding neurons, which connect not only various types of an action (various grasp types), but may also "group" certain actions according to their goal.

3.4 Other possible functions of mirror neurons

Apart from understanding of actions, mirror neurons have also been considered to be involved in understanding of emotions. Gallese et al. (2004) describe the mirror mechanism as a basic functional mechanism that provides an insight into other minds. On the other hand, as recently discussed by Rizzolatti and Sinigaglia (2010), they suggest a dichotomy between motor-mediated and cognitive interpretation of the visual stimuli, emphasizing the superficial nature of the latter type of processing.

Lastly, one of the most intriguing conclusions drawn from the existence of mirror neurons is their possible role as a "missing link" between animal communication and human language (Arbib, 2005). The evidence in favor of this theory was provided already by Rizzolatti et al. (1996) who claim that area F5 and Broca's area might not only be anatomical homologues,[3] but could also share functional properties crucial for development, production and understanding of communication gestures, which gave rise to the evolution of language. Similarly, Rizzolatti and Arbib (1998) state that the way to the evolution of the open vocalization system present in humans (speech) was paved by the evolution of the manual gestural system, facilitated by the action-execution action-observation matching property of neurons in Broca's area.

4 Where do mirror neurons come from?

Heyes (2010) suggests that in the "missing link" account (Rizzolatti and Arbib, 1998) there is an implicit claim that mirror neurons evolved as an adaptation for action understanding. Heyes maintains that according to the *adaptation hypothesis*, experience plays a relatively minor role in the development of mirror neurons (it triggers or facilitates it) and that the capacity of mirror neurons to match observed with executed actions is genetically inherited. She also proposes an alternative, opposing account – the *association hypothesis*, which states that mirror neurons are merely a byproduct of associative learning and that their existence is not caused by any evolutionary mechanisms. Although Heyes argues that these hypotheses are both plausible, she favors the latter providing arguments both for association hypothesis and against the adaptation hypothesis. Regarding the evolution of mirror neurons, Heyes (2010) states that "motor neurons and associative learning did not evolve for the purpose of producing mirror neurons". As another argument against adaptation hypothesis, Heyes uses new experimental results showing that sensorimotor experience can temporarily reverse mirror neuron activations (Catmur et al., 2007). However, we agree with the opposing claim this does not necessarily undermine the role of mirror neurons in action understanding or even their evolutionary benefit (Rizzolatti and Sinigaglia, 2010).

According to the association hypothesis, the motor resonance during action observation occurs due to memory retrieval of the execution of the observed action. The memory, triggered by a visual stimulus, was formed in the past, when the observer executed the particular action with visual guidance. These memory-triggered mirror neurons are a product of associative learning in the sense of Pavlovian conditioning and are extensively trained by correlated experience, even if the executed action is different from a simultaneously observed action. Using this correlation

[3] However, the relationship may be more complicated as suggested by Grezes et al. (2003).

account, Heyes explains differences between humans and monkeys. She claims that "humans receive a great deal of more correlated experience of observing and executing similar actions" and so the human mirror system can react to a greater variety of stimuli. We agree that correlated experience plays a crucial role in the development of mirror neurons. However, we believe they are not an insignificant byproduct of some other processes or a random phenomenon, but a functional piece of a larger mechanism underlying action understanding and (dependent on the location in the brain) understanding in general.

4.1 Reconciliation

We claim that both genetic factors and sensorimotor experience are *crucial* for emergence of mirror neurons. We argue that genetics expresses itself largely in terms of cortical wiring at various levels of granularity, while experience manifests itself in tuning synaptic efficiencies and potentially also in some degree of synaptic rewiring. This nature-nurture dichotomy has been illustrated in various connectionist models of development (Elman et al., 1996). When talking about innateness, one has to specify what is meant by that term. Elman et al. (1996) propose that innateness is imprinted in various constraints operating at three different levels: representational, architectural and chronotopic. Representational constraints are considered very implausible because, in connectionist terms, they would imply prespecified weights of synaptic connections, which in turn would imply inborn representations. Indeed, Elman et al. (1996) provide a mounting evidence against the notion of inborn (domain-specific) microcircuitry (in Chapter 5). This view is hence in clear opposition with classical views of cognition, some of which assume inborn knowledge of concepts, implemented as neural representations.

Architectural constraints are much more credible as supported by numerous neuroscience evidence. These constraints relate to various levels of granularity ranging from neuron-level constraints, such as specification of neuron types and their associated characteristics, via local neural circuits (e.g. layered organization of the cortex at various parts of the brain, degree of interconnectivity in terms of "fan-in" and "fan-out") to global architectural constraints in the brain, like the thalamo-cortical or cortical-cortical pathways.[4] These constraints, however, do not exclude mechanisms for possible rewiring, evoked by changing experience (brain plasticity). Lastly, chronotopic constraints are reflected in the timing of events in the developmental process.

Various stages of cognitive development, for instance in language acquisition, or the "theory of mind", are known to take place at typical age of a maturing child. This may have some genetic basis which can be slightly modulated by individual characteristics and the environment. Therefore, a plausible account for the formation of the mirror neuron system will necessarily depend on obtaining a right balance between nature and nurture factors so that they interact correctly. Oberman and Ramachandran (2008) present a similar view (in an open peer commentary in Hurley, 2008) and in addition, they propose an experiment on newborn monkeys that could help to disentangle the inborn capabilities and learning. They also propose an alternative experiment with adult monkeys that might shed light on whether mirroring ability could be achieved by Hebbian associations.

[4] We argue that it is possible (and necessary) to distinguish between representational constraints and architectural constrains even if they have similar consequences. An instance of a representational constraint can be that a concrete neuron in F5 is connected to the neuron in F1 which is connected to the, say, thumb on the left foot. An instance of an architectural constraint can be the forced growth of neural synapses between F5 neurons and F1 neurons to follow a specific distribution; what in turn causes that some neurons will really (statistically) be connected to the left-foot-thumb neuron in F1.

4.2 Associative versus Hebbian accounts

Heyes (2010) emphasizes the distinction between her association hypothesis and the Hebbian account (Keysers and Perrett, 2004) by claiming that the Hebbian learning only implies contiguity whereas the associative account requires both contiguity – the closer the two events occur in time, the stronger the association, and contingency – required correlation or predictive relationship between the events. We argue that there is no significant difference between the two accounts and that also the Hebbian learning requires both contiguity and contingency, although the latter was not explicitly mentioned in the original Hebb's postulate. This can be demonstrated in the implementation of the Hebbian learning in artificial neural networks, composed of either spiking or artificial neurons. In spiking neural networks, the spike-timing-dependent plasticity (STDP) assumes that presynaptic spikes have to shortly precede postsynaptic spikes (contingency) in order to lead to long-term potentiation (Froemke and Dan, 2002). If the order of spikes is reversed, synaptic depression occurs. At the level of artificial neurons, the same rule applies because the neuron's output is computed (at a discrete time step) as a weighted sum of its inputs which are assumed to be activated at the time of output computation. So the output activation follows input activation (contingency) after which the learning rule can be applied. In addition, there exists a link between the two levels of computational modeling. The Hebbian-like BCM learning rule follows directly from STDP when pre- and postsynaptic neurons are uncorrelated or weakly correlated Poisson spike trains, and when only nearest-neighbor causal spike interactions between neurons are taken into account (Izhikevich and Desai, 2003).[5]

Although the computational mechanisms operate at the level of neurons, it should be kept in mind that the both the associative and the Hebbian accounts of mirror neurons are actually a high-level psychological explanation that links "spatial" (non-sequential) sensory and motor patterns. As argued by Knott (2011) (section 2.7.5.2 of his book), in the context of associations between STS and F5, mediated by PF, the Hebbian account assumes that sensory and motor representations, that have inherently naturally sequential structure, are first independently integrated in time to yield static representations that can then be associated in one-to-one fashion. For instance, a particular sensory representation corresponding to a concrete grasp type is associated with a corresponding underlying motor representation (Keysers and Perrett, 2004). It is an open question whether we can rely on such a type of association, or whether we need to consider the sequential nature of these learned associations.

5 Action understanding as a graded capacity

Heyes (2010), as well as Hickok (2008), suggests that mirror neurons do not play a dominant (if any) role in action understanding. Regarding action understanding, it is necessary to clarify what it means, as authors differ in their positions (for the discussion on this topic see Hickok (2008). We adopt the definition of Gallese et al. (1996) who see action understanding as "the capacity to recognize that an individual is performing an action, to differentiate this action from other analogous to it, and to use this information in order to act appropriately". An important and interesting is the "act appropriately" clause. It implies that the observer has an internal state representing the action and that it is externally possible to assess this state through his

[5] It has been shown that the STDP rule endowed with the BCM sliding threshold can account for homosynaptic LTP accompanied by heterosynaptic LTD experimentally observed in the hippocampus (Benuskova and Abraham, 2007). This supports biological plausibility of such a rule.

behavioral response. Since the correct response should be beneficial to the agent, we can expect and evaluate it in advance. However, the response might be “wrong”, even if the agent believes that he/she is acting appropriately. The actual outcome might differ from the expected one because of some external or internal factors influencing the performance of the agent, but also because of insufficient understanding. The appropriateness of the reaction depends on the *depth* of understanding.

Our claim can be demonstrated on examples of actions whose accurate execution requires certain training. As an example, consider two men watching a football (soccer) match. One of them has played for several years in a university league, whereas the other, although being a keen fan, has never played football himself. Both men observe the same detailed shot of a football player manipulating the ball in front of the goal area in a difficult situation. When commenting on this situation, the unskilled football fan will predict that the player will surely score a goal. However, the skilled observer sees that there is a slight disturbance in the player’s posture and anticipates that the player will fail to score. This difference in the anticipated outcome of the action can be attributed to the depth of action understanding. Obviously the skilled man is able to evaluate the movement of the player, most likely through accurate motor simulation (resonance), but the unskilled man can judge only according to his general knowledge of football and visual memories of it.

This prediction should be testable using a behavioral experiment. At first, one group of participants will learn how to execute a simple, but novel movement, which is not part of their motor repertoire. During this time, the second group will observe videos of the same movement and receive a linguistic description of it. At the end, both groups will have to answer various questions regarding this movement, for example, whether it is possible to grasp an apple with it. Our prediction is that the second group will perform worse than the first group because mere observation of an action prevent the subjects from reaching the same level of motor proficiency, as in the case of subjects who could actively learn the task.

Another study that is in line with our assumption that greater familiarity leads to better understanding, could be seen in the experiment of Knoblich and Flach (2001). Participants were filmed while throwing darts. Then they watched videos of themselves and others throwing darts and were asked to predict the result of the action (each participant was informed whether he was watching himself or another person). The results of the experiment were divided into two subsequent sets, while in the second set participants were expected to perform better, because they had got used to a strange situation of watching themselves. Indeed, participants were more accurate at predicting their actions in the second half of first-person observations and importantly, they were better at predicting the outcome of their own action in comparison with the actions of others.

We argue that action understanding is not a binary but rather a *graded capacity.* An important factor contributing to deeper understanding is, whether the observed action belongs to observer’s own repertoire. Within a repertoire of known (capable) actions, the degree of understanding will depend on the level of observer’s proficiency in that action, especially when it comes to specific actions that require certain training (e.g. kicking a ball in football game, or performing some specific manual actions).

This view of action understanding which should be distinguished from action recognition (see the following subsection) is consistent with the representation of *meaning*, when taking a cognitive semantics perspective. There exist various (categories of) entities that subjects can observe and understand (activate the meaning representation of the observed entity): actions, (static) objects, situations, and concepts. As argued above, action understanding implies the motor simulation at a specific level of accuracy. Since action understanding is a continuum, the observer reacts according to the depth of his understanding.

Understanding an object, rather than merely visually recognizing it, also involves the extraction of object affordances (Gibson, 1977) taking place in parieto-motor areas (in case of object proximity, canonical neurons contribute to this process). We admit that certain understanding of an object (i.e. accessing and processing the conceptual information about it, in order to solve the task) can occur also without the involvement of motor areas (e.g. in case of their dysfunction, as argued in Mahon and Caramazza (2005)). However, also in this case we would argue that deep understanding of an object (i.e. the ability to act upon it directly on indirectly) involves the availability of motor simulation. Similarly, understanding a situation can be argued to imply extraction and mental simulation of affordances pertaining to the situation (i.e. "what I could do in that situation"). Understanding concrete concepts can also be based on sensorimotor experience, in which case these concepts are directly perceptually grounded. Abstract concepts, even though they do not induce direct sensorimotor memories, can be based on perceptually grounded concepts (Barsalou and Wiemer-Hastings, 2005).

5.1 Ventral versus dorsal streams

The question of identification of the neural substrate of action understanding spans also the theory of two streams of visual processing (Ungerleider et al., 1982; Milner and Goodale, 1995). The opponents of the mirror neuron theory, Hickok and Hauser (2010), propose that the substrate for recognition and understanding of actions is localized in the ventral stream, rather than the dorsal stream that mediates communication with motor-mirror areas. As mentioned above, Hickok (2008) emphasizes the role of STS[6] which is localized between the two streams, so we can suppose that Hickok and Hauser (2010), attributing the major role to the ventral stream, also add STS to the core of a single action-understanding object-understanding system[7]. They claim, as well, that the dorsal stream (and mirror neurons) serve only for motor preparation and action selection. On the other hand, the mirror neuron theories attribute the ventral stream also the role of action understanding (see Table 1).

We take into account the dissociation studies (Mahon and Caramazza, 2005) and agree that a system of mirror neurons cannot be solely responsible for action recognition and action understanding. On the other hand, it is important to note that the mirror neuron theories do not exclude the ventral stream, as well as STS, from the action understanding process, since STS, along with IT (a crucial part of ventral stream), project visual information to PFG and to AIP, both endowed with mirror neurons and heavily connected with F5 (see section 2). We maintain that these streams cooperate in order to provide hints and supplementary information

[6] Interestingly, STS has a rather heterogenous structure and probably also functionality, spanning over view-dependent objects recognition, but also view independent recognition, such as view-invariant recognition of faces (Perrett et al., 1991). It reacts also to a large variety of biological movements and to the intentionality of movements (Pelphrey et al., 2004), hence is often thoerized to mediate action-understanding. Despite its structural property, which might divide it to multiple functional areas, STS is often refered to as a single "multi-purpose" entity.

[7] In Hickok and Hauser (2010), "recognition" and "understanding" appear to be the same capacity. In our view they are different. By "recognition" we mean a mere classification/categorization of an object, event or action. Since we consider understanding as a graded phenomenon, recognition would can be equated with shallow understanding.

Table 1: Overview of different positions on mirror neurons. The table describes how the direct-matching hypothesis (Rizzolatti et al., 2001), the position of Hickok and Hauser (2010) and our position explain how the functions of understanding and recognition are split across the dorsal and ventral streams of visual processing.

Account	Dorsal stream	STS	Ventral stream
Direct matching	Object-oriented sensorimotor integration Action-oriented sensorimotor integration Action "understanding"	Object "understanding"	
Hickok & Hauser	Object-oriented sensorimotor integration Action-oriented sensorimotor integration	Object "understanding" Action "understanding"	
Our position	Object-oriented sensorimotor integration Action-oriented sensorimotor integration Action understanding	Object recognition Action recognition	Object recognition

for each other. Similarly in object recognition and understanding, the most recent theories claim that the processing is indeed distributed across the two streams, that encode information differently[8], but are interconnected and depend on each other (Farivar, 2009).

As displayed in Table 1, in accordance with our graded understanding hypothesis, observed action is processed in both streams as well as in STS. Unlike Hickok and Hauser, we differentiate between recognition and understanding and emphasize that understanding involves deeper processing. In the case of hand and mouth actions in the center of mirror neuron theory, the deep processing equals the motor resonance (mirror neuron activity). According to our view, the process initially takes place in STS but in a very short time the information is also projected to F5 (via PF) and back to STS to form a sensorimotor circuit (red arrows in Figure 1) (Matelli and Lupino, 2001). If the corresponding (strictly but also broadly congruent) motor representation is activated, understanding becomes "richer" than mere visual assessment. On the other hand, when the movement is "too far" from the observer's motor repertoire, only the visual analysis in STS provides conclusive information.

Recall the example with a saxophone. Hickok (2008) points out that there is no previous experience with the instrument needed to conclude that someone is playing a saxophone. This is clearly the result of visual analysis in STS. Importantly, STS mediates both object and action recognition, therefore it is likely that a simple recognition of a saxophone as an object is sufficient for categorizing the observed action. Usually understanding is studied as a "binary" phenomenon, one can either understand or not. We argue that also the degree of understanding should be studied to fully assess the brain processes that mediate it. Since understanding is a graded capacity, there is no dividing line between understanding and mere recognition.[9]

Our position is very close to the recent proposal by Tessitore et al. (2010) who claim that all processes of observed behavior draw on the motor area that inherently takes part in the posterior-frontal loop. Therefore, even action recognition involves motor information which, in addition,

[8] The ventral stream serves for general categorization, but also for fine distinction between the members of categories, and most probably encodes objects in an invariant manner. The dorsal stream, on the other hand, is responsible for so called 3-D aspects, as size, shape and orientation of the object, and affects object recognition in the ventral stream. For a conclusive review, see Farivar (2009).

[9] In a wider context, we sympathize with continuous, rather than discrete, nature of the mind and its plethora of processes, of which action understanding is an example (Spivey, 2007).

is assumed to simplify the processing of visual inputs and to improve the action recognition process.[10] We endorse this view that assigns a stronger role to motor areas in socially-oriented behavior, hence providing support for mental simulation theories. Nevertheless, we conjecture that motor areas may not always be a part of the action recognition process, especially when it comes to recognizing actions outside agent's own repertoire. Again, action understanding is a graded capacity, and depends on the degree of motor involvement.

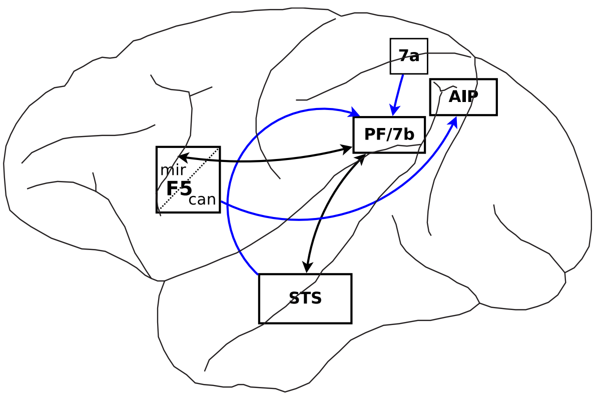

Figure 2: A schematic depiction of the modelled brain areas. F5 is split into mirror and canonical neuron regions for convenience; biolocally they seem to be mixed in the same area.

6 Computational and conceptual models of MNS

Since the discovery of mirror neurons in macaques (and most recently in humans), various conceptual and computational models addressing the MNS have been proposed. Here we focus mostly on computational models that provide concrete processing and learning mechanisms of the MNS, as well as generate predictions. Most of these models are naturally implemented by artificial neural networks based on various learning paradigms. An earlier overview focused on several models that attempted to account for imitation (Oztop et al., 2006).

6.1 FARS model

The very first modeling effort that is related to MNS is the FARS model (Fagg and Arbib, 1998). Rather than directly approaching the mirror properties, the model focuses on visually guided grasping of objects and analyzes how the canonical part of the MNS circuitry, centered on the AIP→$F5_{\text{can}}$ pathway in macaque, may account for this ability (see Figure 2); more details can be found in Fagg (1996). In the computational model, AIP represents the grasps afforded by the object while $F5_{\text{can}}$ selects and drives the execution of the grasp schemas. Hence, AIP and $F5_{\text{can}}$ may act as part of the visuomotor transformation circuit, coupled with a modulating mechanism for affordance extraction via ventral pathway (through IT and PFC→AIP).

In the FARS model, this visuomotor transformation is heuristically hardwired but, as an interesting and important feature of the model, it was shown that the transformation can be

[10] Actually, Tessitore et al. (2010) only talk about action recognition but since they argue for the support of the direct-matching hypothesis, we can assume that they also have action understanding in mind.

learned using the reinforcement learning (RL) in which the reward signal provides feedback about the success and efficiency of the chosen grasp action and guides parameter tuning of the neural network for better grasp configuration (Fagg, 1996).

6.2 MNS1 model

The MNS1 model (Oztop and Arbib, 2002) extends the modeling perspective by assuming the actor and the observer of an action, hence embracing the MNS system. The model inherits the conceptual basis of the FARS model but augments it by mechanisms that can recognize an action in terms of the *hand state*, a novel representational element, which makes explicit the relation between the unfolding trajectory of a hand and the affordances of an object. The hand-state representation includes not only hand features but also hand-relative position with respect to the target object (assumed to merge in area 7a) that both predict specific grasps. Oztop and Arbib (2002) focus attention on the relation between 7b, AIP and $F5_{\mathrm{mir}}$. They show that a feedforward two-layer neural network (NN) with perceptron units, representing a 7b→$F5_{\mathrm{mir}}$ mapping (a core mirror circuit), could be trained (by the error back-propagation algorithm) to recognize the grasp type from the hand-state trajectory, often achieving correct classification well before the hand reaches the object. The NN takes as inputs the hand state representation that in the brain is probably formed in STS (hand shape recognition) and 7a (hand/object spatial relation), and generates as outputs action recognition signals ($F5_{\mathrm{mir}}$ layer). The hidden layer of the NN (layer 7b) that mediates the mapping corresponds to the object affordance-hand state association detectors. The model was simulated in a scenario that allows to generate training sequences as well as target responses. The activity of the $F5_{\mathrm{can}}$ neurons serves as a teaching signal for the $F5_{\mathrm{mir}}$ neurons to enable the monkey to learn which hand-object trajectories correspond to the canonically encoded grasps. As a result, the appropriate mirror neurons come to fire in response to viewing the appropriate trajectories even when the trajectory is not accompanied by $F5_{\mathrm{can}}$ firing. The information provided by the hand state is preprocessed, using an object-centered frame of reference, to yield an invariant representation (with respect to the agent of the action), allowing action recognition (grasp type), manifested by $F5_{\mathrm{mir}}$ firing. This enables the 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.

The MNS model has nice functionality, yet it is based on two questionable assumptions: (1) that actions being learned (by $F5_{\mathrm{mir}}$ neurons) are already in the monkey's repertoire, and (2) that the hand-state trajectory can be converted to a spatial representation to serve as an input for the NN. Assuming that $F5_{\mathrm{can}}$ are trained before $F5_{\mathrm{mir}}$ implies that the monkey first learns to perform and recognize its own actions (and object affordances) that are not yet registered by its $F5_{\mathrm{mir}}$ neurons, and only afterwards this knowledge is learned by $F5_{\mathrm{mir}}$ for monkey's own actions as well as observed actions. Generating invariant hand-state representations relies on the availability of this information provided by STS (Olson, 2003). This high-level visual information is hence calculated outside the mirroring system but we think that feature extraction is crucial for MNS development and should form a part of the model.

6.3 RNNPB model

By taking a dynamic systems perspective, the model of Tani et al. (2004) represents a shift towards recurrent architectures, which allow learning sequences. The recurrent neural network with parametric biases (RNNPB) was designed to allow learning, imitation and autonomous sequence generation. The crucial novel feature of RNNPB are parametric biases (PB) – input units with adaptable activations,[11] that are associated with the spatio-temporal patterns, and are considered analogous to mirror neurons due to their encoding property. The model with modified Jordan architecture (with one hidden layer) takes the current sensorimotor feature vector as input together with PB units and current context representation that has a delayed feedback to itself (see Figure 3). The output units aim to predict the sensorimotor activation vector at the next time step. All layers except the input layer have sigmoid activation functions.

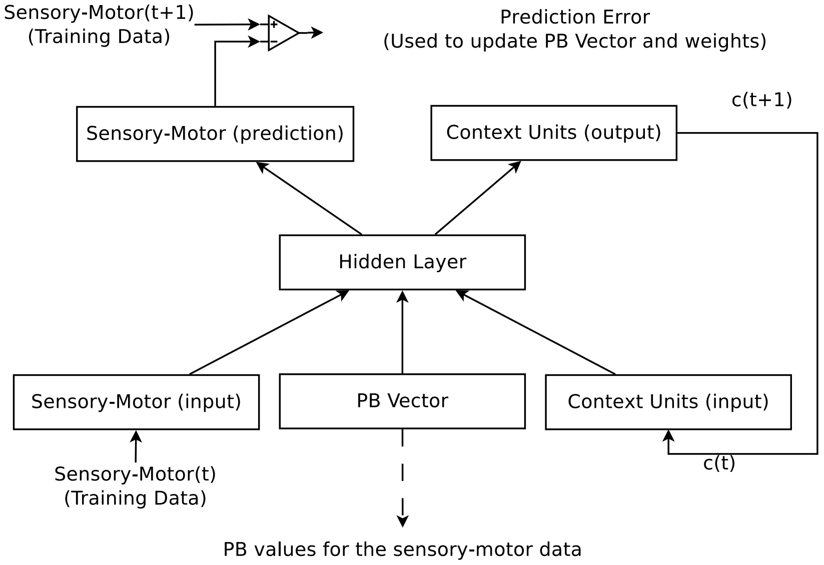

Figure 3: The architecture of the RNNPB network that can operate in three modes: learning mode, recognition mode and generation mode. The exact architecture of the model and the information flow is mode-dependent. For explanation see the text.

The RNNPB model operates in three modes: learning mode, recognition mode and generation mode. In the learning mode, the pre-generated sensorimotor sequences are presented as inputs (one at a time) to the network trained as the next pattern predictor by supervised BPTT learning algorithm (Rumelhart et al., 1986). All network connection weights are adapted and so are the PB unit activations (that are randomly initialized at the start) in analogical way. As a result of this update, each PB activation vector becomes associated with a particular input sequence. In the recognition mode, only the PB vectors are updated but not the weights of the network. The task of the network in this mode is to observe an ongoing behavior (sensory inputs only) and compute a PB vector according to the observed behavior. The error between the predicted next sensory input and the actual next sensory input is back-propagated to the PB vector to drive its update (such that the prediction error is reduced). In this mode, the feedback is restricted only to the motor component. In the generation mode, the network is (externally) set a PB vector and is expected to generate the corresponding, previously learned

[11] The method of training the input units was inspired by the original application to sentence processing in Miikkulainen and Dyer (1988). Learning the representations of input units is actually implemented as an extension of the error backpropagation algorithm.

sequence. No learning takes place in this mode.

The RNNPB model should lend itself to generalizing to novel sequences which is a crucial property of any learning system. Indeed, Cuijpers et al. (2009) experimented with the RNNPB model and showed that the network is capable of recognizing noisy action sequences and of generalizing from a few learned examples. First, they added noise to learned sequences and observed to what extent the network was able to correctly generate or recognize the sequence. Second, they systematically modified the input sequence parameters (frequency and the amplitude of sinusoidal patterns) and evaluated the changes in corresponding PB vectors as a result of this modification. The simulations revealed that the recognition of action sequences was quite robust against noise. They found that the RNNPB network generalization to the frequency was good but it was limited with respect to the amplitude.

Tani et al. (2004) do not attempt to relate the RNNPB model to anatomical regions in the brain, except saying that PB units could serve as analogues of mirror neurons. However, the input (and output) sensorimotor layer of the model spans both temporal (probably STS) and frontal areas of the cortex, and consequently does the context part of the developing representations. Such a layout differs from other attempts such as MNS1 (where network layers represent brain areas) but could probably also be feasible to defend.

It is important to note that the relationship between the PB vectors and mirror neurons in F5 seems only as a weak parallel. Although the PB vectors, like mirror neurons, both characterize and trigger various sensorimotor sequences, their nature is rather amodal than motoric. Since the input layer can consist of various inputs, here of visual and motor information, the PB vectors can bind information and produce estimates in any modalities depending on the user's interpretation of the input layer (where motor and visual inputs are concatenated in the form of numbers). On the other hand, the activity of mirror neurons can be triggered solely by visual input, but their major role is in motor control. In this sense, PB vectors are more likely to be some pointers in memory that can trigger actions. Along with the recent evidence on mirror neurons in humans (see section 2.2), they could be localized in hippocampus or neighboring structures (Mukamel et al., 2010), which are also endowed with mirror properties, but serve like memory pointers (O'Reilly and Munakata, 2000, p. 289).

As for plausibility from a developmental perspective, the RNNPB model rests on an assumption that can hardly be met in ecologically valid scenarios. The learning regime requires that target sensorimotor trajectories (that would allow imitation) are available prior to training. Tani et al. (2004) obtain these trajectories using an optical engineering method, by mapping the (observed) user's arm position to the robot joint angles (via an intermediate stage, robot's arm 3D positions in robot-centered frame of reference) by solving the inverse kinematics problem. In real setting, however, the required joint angles would not be available to the agent observing user's actions. Actually, this mapping would correspond to the developmental stage of an agent who has already learned its own sensorimotor correspondences and now faces the task of mapping observed sensory sequences to its own motor sequences. It is also not clear to what extent the model would generalize if we considered the wide spectrum of angles of observed actions of the same type that should all trigger the same PB pattern. (The invariance issue was also addressed in the case of MNS1 model.)

6.4 MSI model

Oztop et al. (2005) developed a computational model of mental state inference by "putting

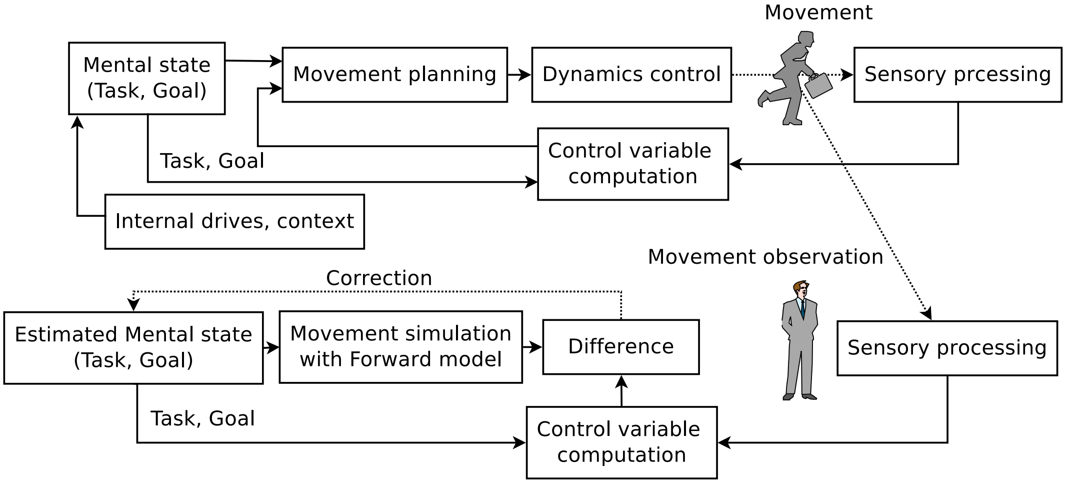

Figure 4: MSI model in action execution mode (upper part) and action observation mode (lower part). The observer tries, using its own sensorimotor loop, to infer actor's mental state. If successful, the observer's simulation loop represents understanding of the observed action. (Adapted from Oztop et al., 2005).

oneself into the actor's shoes" approach with a purpose to account for understanding the observed actions. The functionality of the model is investigated in the context of grasping actions (see Figure 4). As a first step, the model learns to grasp an object using a visual feedback, exploiting both inverse and forward models.[12] Specifically, the parietal cortex (probably area 7b) extracts visual features from the visual cortex – the control variables, relevant in the execution of a particular goal-directed action. These variables could include, for instance, the distance between the hand and the object, in case of reaching. The premotor cortex (probably $\mathrm{F5}_{\mathrm{mir}}$) receives this information and computes the motor signals (inverse model) to match the parietal cortex output to the desired neural code relayed by the prefrontal cortex. The premotor cortex also includes the forward model that mediates inverse model learning for establishing a feedforward control strategy.

In the MSI model, the function of the forward model is extended beyond motor control, as it is used in simulating the mental state of the observed actor and hence, decoupling the mental movement execution from the actual one. So, once the agent's own sensorimotor feedback loop is established, its use is extended to observed grasping behavior. In other words, the MSI model involves a "mental simulation loop", built around a forward model, which in turn is used by a "mental state inference loop" to estimate the goal of the observed agent.

The mental state inference loop is trained by a gradient-based method that minimizes the error between the predicted sensory outcome generated by the observer's forward model and the observed sensory outcome. If the error is minimized, the observer can infer actor's mental state and hence accurately predict the trajectory of the observed action.

The crucial assumption, necessary for functioning of the model, is the use of object-centered

[12] Inverse and forward models are canonical components in the motor control theory but are also assumed to serve as internal models of the brain. An inverse model transforms a desired sensory state into a motor command that can lead to it, whereas a forward model predicts the next sensory state (consequence) resulting from the motor command (Kawato, 1999).

frame of reference for both executed and observed grasping movements (extracted control variables are invariant under translation and rotation), attributable to STS, which allows the inference of actor's mental state.

The MSI model suggests that action understanding requires mental (motor) simulation of the observed action, what shifts the MSI model closer to the direct-matching hypothesis. We believe that the motor simulation associated with perspective change occurs in many cases, which require conscious effort to take allocentric perspective. There is even a recent behavioral evidence by Frischen et al. (2009) showing that change in perspective taking can occur subconsciously. Their research suggests that witnessing another's action leads the observer to simulate the allocentric selective attention mechanisms such that they effectively perceive their surroundings from the other person's perspective. The question is whether understanding the grasping movements also requires a change in perspective[13] taking or whether this ability can be achieved without it, as in the case of predicting the action trajectory by merely extrapolating the visual information.

6.5 MNS2 model

The MNS2 model of Bonaiuto et al. (2007) is an extension of MNS1 both in terms of architecture and representations. It is based on the modified Jordan recurrent network that closely resembles the RNNPB model, and is trained by BPTT algorithm, to be able to classify three types of simulated trajectories that represent different object grasps. The two-layer network receives as inputs hand state representations (mentioned in section 6.2), extracted by means of using the object-centered frame of reference, and learns to predict the trajectory type for a fixed target. The predicted values include the categorical information that was used during training, and an internal signal that serves as a context for the network in the next step, together with the state information. The authors interpret the activity of the output neurons as an analogy of mirror neurons, and the target activity as the activity of canonical neurons.

The hidden layer is assumed to correspond to area 7b that performs the object-affordance–hand-state association. The input to this layer comes (in the form of a 7-dimensional vector as in MNS) from areas 7a and STS. Hence, the MNS2 model is designed to perform (STS+7a)→7b→$\mathrm{F5_{mir}}$ mapping in which the internal representations on the hidden layer emerge as a result of the required input-output mapping.

The important component added to the model is the working memory for the hand and the object, that predicts correct trajectories (by extrapolation) even when the hand becomes temporarily invisible. Another nice feature is the addition of an auditory subsystem (with the same architecture) and the resulting formation of multimodal (audio-visual) representations of mirror neurons (using Hebbian associations) between actions and their characteristic noises (such as peanut breaking or paper ripping). This way, the mirror neurons learn to respond not only to visual but also to associated auditory stimuli.

The MNS2 model shares with MNS1 two assumptions: the use of an object-centered frame of reference and the development of $\mathrm{F5_{can}}$ neurons, taking place before training $\mathrm{F5_{mir}}$ neurons, so that they could serve as targets. Related to the second assumption, this learning should involve mechanisms that would control the activation of canonical neurons (e.g. by inhibiting them) during action observation, when they are known not to fire.

[13] There is a difference between the following three approaches: direct coordinate translation (computationally intensive and therefore questionable), change of attention (Frischen et al., 2009), and view-independent object recognition in macaque STS (Perrett et al., 1991).

6.6 Extension of MNS2 model

Bonaiuto and Arbib (2010) came with an idea that the mirror neuron system might mediate another function, which they called "what did I just do" function. The motivation came from experiments with a cat which, after spinal lesions affecting grasping with a forepaw, re-learned to extract food from a horizontal tube in a different way than an unimpaired cat (Alstermark et al., 1981). To simulate this, Bonaiuto and Arbib (2010) designed an extended architecture of the MNS2 model which uses the MNS as a component in a feedback loop that updates weights in the model using a RL method (which they named augmented competitive queuing).

The model uses representations of the external state (distances between important objects) and the internal state (hunger) as inputs into two separate systems: the actor system and the mirror system. The actor system chooses an action based on its desirability, computed from the internal state, and its executability, computed from the external state. The model assumes a limited repertoire of meaningful actions (such as grasp-with-paw, grasp-with-jaws, rake, raise/lower neck etc.) and a variable repertoire of meaningless, "irrelevant" actions, to enlarge the search space for useful actions after the lesion. The final action is chosen according to its highest priority, computed as a product of its executability and desirability.

The learning in the model is, in comparison with the earlier models, quite complex. What we find interesting about the model is its reliance on RL paradigm which is considered to be ecologically valid. RL is applied after the recognition of action for updating both executability and desirability of the action. Reinforcement signal for the executability is positive (action is reinforced) if it was recognized by the MNS as a successful action. It is negative if its MNS activation was below a minimal level, for instance when action was not recognized as successfully performed, but it was the intended action. Otherwise, executability is not affected. The executability is thus reinforced by the perceived ability to perform the recognized or the intended action.

RL signal for desirability is formed by the presence of food in mouth and processed by the adaptive critic module that compares predicted desirability of recognized action with the presence of primary reinforcement (food in mouth). Effective RL signal is computed as a sum of primary reinforcement and the error in current prediction of desirability, being a difference between discounted predicted desirability of the current action and the predicted desirability of the previous action. Thus, if the next action is more desirable than the previous one, the RL signal for recognized action is slightly positive, as it predictably brings the cat closer to the goal. If the food ends in the cat's mouth, the RL signal is maximal and the desirability of the just-performed action is increased strongly, later to serve in reinforcement of actions leading to it.

Once the agent learns to perform all actions well, the lesion is induced in form of noise and inaccuracies at the input and the agent's motor schemas, and the desirability and the executability of actions have to reorganize. If the MNS is left out, the reinforcement is markedly slower because the agent has no clue as to what action was actually performed, only the primary reinforcement (food) is available. However, if present, MNS is shown to mediate rapid reorganization of successful behavior to compensate for the lesion.

The extension of the MNS2 model highlights the fast reorganization ability of the MNS by approaching it in a cognitively plausible way (using RL). The proposed computational mechanisms are therefore very interesting.

6.7 MOSAIC

The MOSAIC model is a rather sophisticated architecture that was originally designed for motor control (Haruno et al., 2001). However, it was shown to be also usable for mirroring purposes, such as action recognition and imitation. MOSAIC is modular and allows a distributed cooperation and competition of the internal models (see Figure 5). The basic functional units of MOSAIC are multiple predictor–controller (i.e. forward–inverse model) pairs with each competing to contribute to the output, such that controllers with better predicting forward models become more influential to the overall control (hence, MOSAIC stands for modular selection and identification for control).

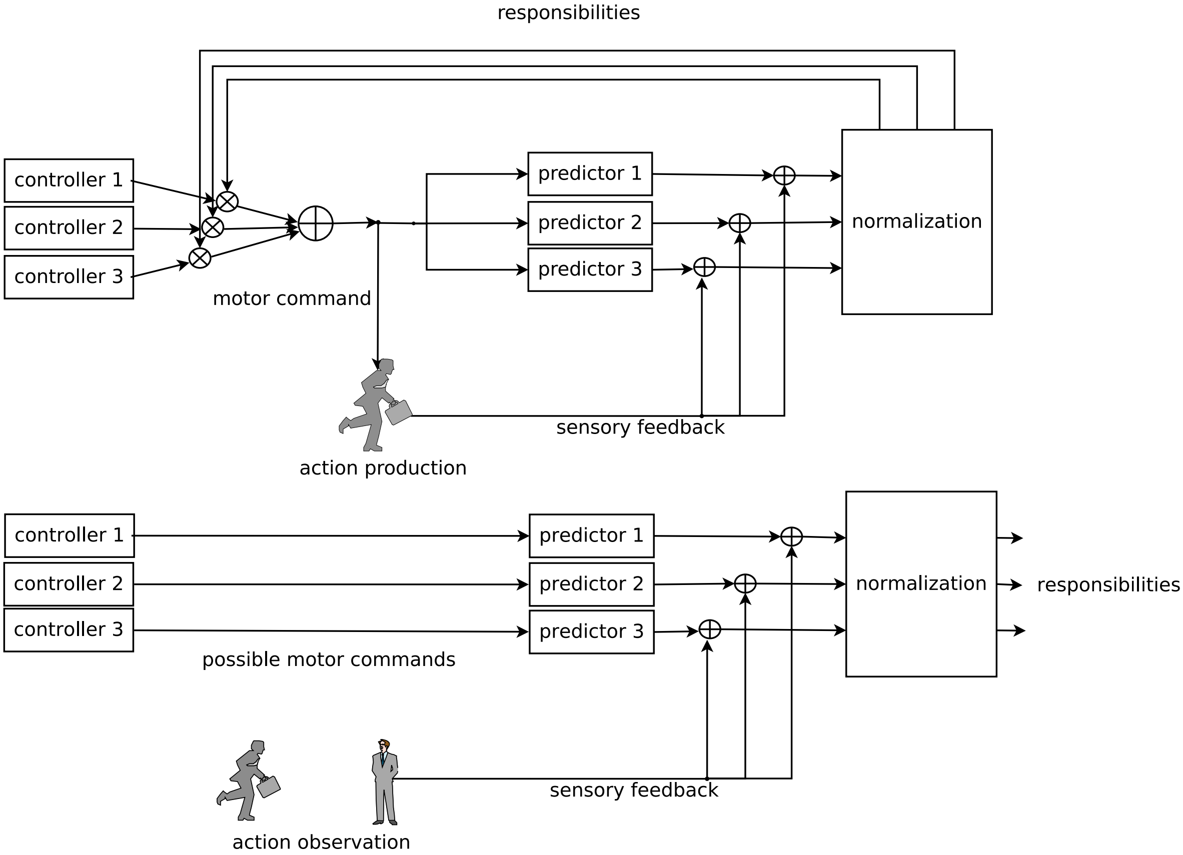

Figure 5: MOSAIC model in action control mode (upper part) and in action observation mode (lower part). For explanation, see the text. (Adapted from Wolpert et al., 2003).

The use of MOSAIC for mirroring in case of action recognition or imitation requires three stages (Wolpert et al., 2003): First, the visual information of the actors movement must be converted into a format that can be used as inputs to the motor system of the observer. This conversion requires that the visual processing system extracts variables related to the state (e.g. joint angles) which can be fed to the observer's MOSAIC as the "desired state" of the actor. The second stage is that each controller generates the motor command which is assumed to achieve the observed trajectory (i.e. the desired trajectory obtained from the observation). In this "observation mode", no movement generation occurs but the outputs of the controllers are used as an input to the predictors paired with the controllers. Hence, the outputs of the forward predictions are the next likely observer's states. These predictions can then be compared with the actor's actual next state. The difference – prediction error – indicates, which of the controller modules of the imitator must be active to generate the observed movement. The outcome of this stage is a symbolic sequence composed of indexes of the controllers selected (one at a time) during action observation. In the third stage, this sequence can either be compared in memory (action recognition) or used for repeating the action (imitation).

The output of predictors might be considered analogous to the mirror neuron activity. The output of the forward model in MOSAIC is related to the intrinsic variables of the controlled limb, like joint positions and velocities, and contrasts with the MSI model, in which it is related to visual-like coordinates, such as the orientation difference of the hand axis and the target object (Oztop et al., 2006).

The distribution of control in the MOSAIC model is an interesting idea that allows efficient control and sounds plausible, given that the brain also probably uses multiple controllers specialized for various behaviors (e.g. grasp types). However, the explicit specification of the number of controllers in advance seems less plausible. These modules and their division of labor should instead be formed during experience. Another difficulty arises within the first stage – extraction of proprioceptive information from the visual input, which is the same problem that we encountered in the case of the RNNPB model.

6.8 Higher-order Hopfield network

Chaminade et al. (2008) designed and implemented an alternative recurrent neural network model (HHOC) that learns visuomotor associations in a robotic hand, designed to account for imitation capacity originating from self-observation. Hopfield's associative memory, being a one-layer recurrent network with full connectivity and symmetric weights, is known to be able to store patterns and, as a crucial property, also retrieve the pattern, given a partial representation of it (Hopfield, 1982). The HHOC model, as an extension of Hopfield network, was used to associate the visual signals presented on the retina and the motor signals that were both generated during the "motor babbling" stage. The Hebbian-like pattern association of (static) signals is achieved in the HHOC model in a somewhat more complex way (cf. Hebbian account by Keysers and Perrett, 2004 in section 6.9), in which each visuomotor pattern is stored as an attractor with its surrounding basin of attraction that enables convergence to the attractor (representing a pattern-recall process). Specifically, the artificial input patterns consist of all possible hand postures with 4 fingers (all but thumb) up or down, and the expected retinal images for the posture coded by the motor patterns. Notably, due to high overlap between (visuomotor) patterns to be stored,[14] the HHOC uses the second-order units (rather than standard units) whose synaptic weight changes are proportional to a product of three unit activations (rather than two). The HHOC model was shown to be quite robust to noise, to generalize across patterns by inducing a correct motor pattern when cued by a novel visual representation, and to generalize across agents by inducing an expected motor pattern when processing a visual input generated by a different artificial/human hand. The authors interpret the latter ability as imitation, because it leads to correctly evoked motor pattern.

In the HHOC model, the visual and motor patterns are mediated by induced attractors. The availability of both pattern types can be safely assumed, if one focuses on motor babbling rather

[14] that would cause disturbing interference during recall

than object reaching/grasping, in which case the motor patterns have to be learned. Testing for generalization to other agents resides in using altered visual stimuli (hands) that evoke slightly different retinal images to be processed. In other models, generalization of this type is related to a perspective change, assumed to have been solved by STS yielding object-centered (rather than viewer-centered) representations. The use of the second-order units in the HHOC model, which leads to higher computational power, solves the modeling problem, and has also been argued to be biologically plausible (Mel and Koch, 1990). The HHOC model has not been linked to any anatomical areas but, in the context given above, we can assume that it is meant to link STS with $F5_{\mathrm{mir}}$ in the form of long-range attractors, ignoring the mediating stages of processing performed by the parietal areas.

6.9 Hebbian account

Keysers and Perrett (2004) provide a Hebbian account of the MNS emergence and function. They focus on $\mathrm{STS}\rightarrow\mathrm{PF}\rightarrow F5_{\mathrm{mir}}$ network in which the mirroring of neurons in PF and $F5_{\mathrm{mir}}$ is hypothesized to emerge as a direct consequence of the anatomical connections between STS, PF and $F5_{\mathrm{mir}}$. The connections between these areas become first associated during self-observation while executing an action, so the activations in STS neurons responding to the sight of this action (e.g. precision grasp) overlap in time with activity in the PF and F5 neurons that fire when the agent performs that action, creating a prerequisite for Hebbian learning. The same logic applies to learning other's actions. Thanks to invariant properties of (object-centered) STS neurons, the observation of someone else performing a similar action will then also activate the same neurons in PF and F5. This way, mirror properties will emerge. The authors also explain how in a similar way the mirroring of unseen actions can emerge (e.g. one's own mouth movements) and how many other forms of social learning can be conceived with Hebbian learning. In addition, this conceptual model includes the explanation for mechanisms that could inhibit STS activations during action execution.

The Hebbian view is clearly a well understandable and relatively simple account for emergence of the MNS. From the computational perspective it may be thought of as a higher-level account because it abstracts away from the sequential nature of sensory and motor signals. Hence, it associates activations at two different locations. Despite its psychological plausibility it is not clear whether this is neurally the right level of association and whether we should not instead associate sequences (see the comment in section 4.2). Sequence association has been approached in the two models (MSI and RNNPB) mentioned above, using different learning paradigms. In addition, it is not clear in the Hebbian view, why only a subset of biological neurons learns the associations to become mirror neurons, whereas other neurons in the same areas do not (Oberman and Ramachandran, 2008). Probably some (lateral) competitive mechanisms coupled with self-organization need to be included to provide an explanation.[15]

6.10 Knott's model

As a part of his theory about a tight link between sensorimotor cognition and natural language syntax, Knott (2011) proposes a quite elaborated (albeit not implemented) model of the mirror neuron circuit (section 2.7.5 in his book), that combines the ideas of the Hebbian account

[15] A theoretical option is that some neurons could be predetermined to become mirror neurons, but this could be considered a rather strong nativist assumption.

(Iacoboni et al., 2001; Keysers and Perrett, 2004) and the forward modeling (Oztop and Arbib, 2002; Oztop et al., 2005). The model focuses on the STS-PF-F5 circuit whose parts are assumed to learn to bidirectionally associate visual representations in STS with motor representations in F5. As a core assumption, shared with earlier models, STS can form action representations, invariant with respect to the agent. As an important part, the inclusion of the forward model aims at capturing the temporal nature of the signals to be associated (see also our comment in the previous section) as well as simplifying the matching to be learned between F5 and STS. The forward model F5→PF→STS converts motor signals into anticipated sensory consequences (in line with Miall, 2003), which may be easier to match with STS signals than the motor representations from which they derive (because they are in the same domain).

Knott (2011) points to several significant differences between his model and several earlier models. The model of Oztop and Arbib (2002) assumes that STS is directly involved during the execution of reach/grasp actions, while in Knott's model it is not.[16] The model of Oztop et al. (2005) differs from Knott's model in two aspects. First, representations of the observed agent's hand are primarily computed in the parietal area, and then are sent to STS for matching with hypothesized representations, which reduces the role of STS. In Knott's model, STS receives inputs from the grasp/reach pathway (MT and MST) but also from the form classification pathway, making STS more autonomous in generating visual representations of actions (analogical to object classification pathway). The second difference is that in the model of Oztop et al. (2005), the learned association between F5 and STS only runs in one direction, from F5 to STS, even during action recognition. This in fact points to the above mentioned reduced role of STS in action recognition. On ther other hand, Knott's model assumes that STS generates its own invariant visual representations that can trigger activity in F5, hence making the links bidirectional.

Knott's model is definitely worth implementing (using neural networks) such that its functionality and predictions could be better appreciated.

6.11 Tessitore et al.'s model

The very recent computational model of Tessitore et al. (2010) differs significantly from all preceding models by involving mirror neurons in coding and making the motor information available for both action execution and action observation processes. As the authors argue, earlier models "identify mirror neuron activity in action observation with the final outcome of computational processes unidirectionally flowing from sensory (and usually visual) systems to motor systems". On the contrary, modelling action recognition by the visuo-motor loop is emphasized to be a distinctive feature of this model, making it coherent with the direct-matching hypothesis. Tessitore et al. assign a central functional role to mirror mechanisms in visual perception, predicting that motor information coded by mirror neurons simplifies the processing of sensory (visual) inputs and improves the results of action recognition tasks.

While focusing on grasping actions, Tessitore et al. (2010) address two questions regarding the mirror neuron activity: *what* it codes, and *how* it bears on sensory processing. The answer to the *what* question is based on the suggestion that hand-joint configurations for grasping actions can be expressed by a small set of parameters (Mason et al., 2001). Tessitore et al. propose

[16] However, the STS 'match' region, discovered by Iacoboni et al. (2001) to be involved in both execution and observation, is active in Knott's model during action recognition, as it receives signals from the motion recognition pathway, namely MT and MST.

that the set of object-directed actions can be subdivided into distinct classes, each one of which is identified by means of a small set of vectors spanning an action subspace in the space of hand-joints configurations. Thus, each hand-configuration can be expressed by suitably setting coefficients in a linear combination of these vectors. The functional role of the mirror neuron activity during action execution and action observation is identified in the model with an action subspace selection process.

This also leads to the answer to the *how* question. The traditional hypothesis, according to which action recognition is based on the mapping from sensory (mostly visual) input to a view-independent action representation, is replaced by transforming the sensory input into intrinsic action features, simplified on the basis of motor information expressed in terms of action subspaces. In other words, the sensory input is mapped on each grasping action subspace, codified in the motor system, using a set of specialized submappings, with a subspace selected by mirror neuron activity. In sum, action recognition process truly involves bidirectional information flow, with a central role played by mirror neurons.

Tessitore et al. (2010) compare their model with the MOSAIC and RNNPB models, that also allow for iterative interactions between sensory processing and mirror activity. However, in these models where the iterative interaction occurs via coupled forward-inverse models, the nature of mirror neurons is claimed to be different. The sensorimotor loop modelled in MOSAIC and RNNPB involves a kinematic (sequential) level of description for actions. On the other hand, in Tessitore et al.'s approach, the motor information encodes action subspace probabilities which refer as a whole to action classes, without requiring a precise reference to action kinematic parameters. This means that involvement of motor information operates on a higher (category) level when interacting with sensory processes, which resembles the level of Hebbian account. The supporting fact for this level of representation comes from the evidence that $\mathrm{F5}_{\mathrm{mir}}$ neurons respond largely unselectively to the kinematic characteristics of executed/observed actions (Craighero et al., 2002).

6.12 Other models

Finally, let us briefly mention two other models, based on different formal methodologies, that widen the spectrum of computational approaches. Wiedermann (2003) proposes the conceptual model of the MNS by taking the finite-state machine perspective. He defines the finite cognitive agent (FGA) that consists of perceptional-motor units (PMU) representing agent's body and a finite-state transducer (FST), representing the mind/brain. FST transforms sensory and proprioceptive data into motor data that are sent back to agent's motor PMUs (these represent multi-modal information by concatenating sensory, proprioceptive and motor information). FGA can function in a standard mode, performing an action with self-observation, and in an observational mode, when it observes an action performed by another agent. The functionality of the agent is explained in quite a detail, but the model was not implemented, rendering the specification somewhat ambiguous.

The purpose of the model was to abstain from technical details and focus on general framework of using the mirroring mechanism to create a valid model of agent's world and basic language skills. Therefore this model does not describe how an agent could acquire the mirroring properties.

Kilner et al. (2007) take a Bayesian perspective to the MNS and identify a precise role for the MNS in the agent's ability to infer intentions from the observed movement. They focus on an (ill-posed) problem of inferring the cause of an observed action and suggest that the problem could be solved by the MNS using predictive coding on the basis of a statistical approach known as empirical Bayesian inference. This means that the most likely cause of an observed movement can be inferred by minimizing the prediction error during action observation by comparing the predicted the kinematics on the basis of agent's own action system and the observed kinematics. The computational mechanisms for this generative approach (as opposed to one-way, recognition approaches) are proposed to exist in two directions. Posterior-frontal pathway STS→PF→F5$_{\mathrm{mir}}$ implements the prediction error of the motor action and frontal-posterior pathway STS←PF←F5$_{\mathrm{mir}}$ implements the generative model that computes the sensory consequences of the performed action (forward model). The authors suggest that the forward model be an integral part of the motor function, since the same model becomes exploited both for action execution and action observation (similarly to the MSI model).

The nice features of the Bayesian framework are that it is principled and exploits bidirectional connectivity. It is consistent with the theory that the brain performs all kinds of predictions (expectations) at various levels of its organization (Friston, 2003). On the other hand, it could be argued that probabilistic modelling is a higher-level account that does not provide neurally-inspired learning mechanisms that are characteristic of connectionist approaches.

7 Discussion

The discovery of mirror neurons has triggered a great deal of empirical, theoretical and computational research. Yet, when considering the variety of papers published in the last decade, the function of the mirror neurons seems to remain an open question, especially when it comes to their functionality in humans. Along with Oztop et al. (2006) we hold the view that the function of the mirror neurons must be rooted in motor control, which is from the evolutionary perspective a very old capability. The mirror function may work for explaining action understanding and/or imitation. Nevertheless, it is unclear whether all hypothesized mirror functions can be explained using the control-theoretic framework (as suggested, e.g. by Hurley, 2008) because there is no evidence that motor control circuits could account also for understanding emotions or mind-reading (Goldman, 2008). Therefore, the function of mirror neurons might be different for different parts of the brain. The most important aspect is the perceptually triggered resonance of mirror neurons that provides a simple mechanism for mental simulation of the observed behavior, which leads to its understanding.

7.1 The role of STS

In case of action understanding and recognition, which is a central theme of this paper, the role of STS has probably not yet been fully appreciated. Neurons in STS do not have mirroring properties, yet a small fraction of them provides highly abstract, viewer-independent representations that may simplify the link to the motor representations (Perrett et al., 1991). These STS cells code information in object-centered rather than viewer-centered frame of reference (Olson, 2003). As an important assumption, the existence of viewer-independent object representations was utilized in several computational models that could take advantage of the agent-invariant visual patterns used for both executed and observed actions, hence greatly simplifying the mirroring task.

It has been proposed that cells responsive to many views of an object or action could be established by combining the outputs of several view-sensitive cells (scattered in the same anatomical area) tuned to different views of the same object or action, which plausible also from the modeling viewpoint. Indeed, such a scheme of processing has been suggested many times on the basis of neurophysiological evidence (Perrett et al., 1991; Hasselmo et al., 1989). An analysis of the time course of responses seems to support this account, since view-selective cells respond at a slightly shorter latency compared to view-invariant cells (Perrett et al., 1991).

It is interesting that the difference between view-dependent and view-independent responses in STS cells is not strict but rather graded. That is, some cells display only partial invariance, thereby possibly serving as an intermediate stage of hierarchical processing. This seems analogous to the congruency of mirror neurons found in the frontal cortex. From the learning perspective, the varying degree of response selectivity determines the degree of generalization when reacting to an input pattern and represents a ubiquitous feature in pattern recognition.

On the other hand, the reliance on the invariance assumption was questioned by Tessitore et al. (2010) who argue that motor areas must be involved in the process of interpretation of visual inputs, contributing to action recognition/understanding, and not only as recipients of an already preprocessed visual information. They support their argumentation by simulating the classification of reach-to-grasp actions, where "mirror-coded information was found to simplify the processing of visual inputs and to improve action recognition results with respect to recognition procedures that are solely based on visual processing" (Tessitore et al., 2010). This theory is also supported by neuroanatomy because it is known that $F5_{\mathrm{mir}}$ area is reciprocally connected to PF, and through PF to STS (Matelli and Lupino, 2001).

It remains to be determined how the information processing in STS providing the viewer-invariant representations is achieved: whether it "merely" includes a bottom-up ventral pathway (maybe taking information from MT/MST areas, as suggested by Knott, 2011), or whether it necessarily involves motor information mediated by the parietal area. We have argued that STS is involved in robust action recognition which in itself is not assumed to require the motor component. Indeed, humans are capable of recognizing (categorizing) a wide spectrum of biological (and non-biological) actions, although some of these cannot be mapped to their own motor repertoire. Action recognition, as we define it, may be achievable by mere visual inspection and categorization of the sequential patterns. In addition, it appears that STS neurons with goal-coding properties (area TEa) provide the visual system with a rich understanding of the world which embodies causation and intentionality (if contextual information is included; van Rooij et al., 2008), without necessarily involving the frontal areas (Perrett et al., 1991).

7.2 Graded action understanding

On the other hand, we argue that action understanding requires "stepping into the actor's shoes." Action understanding is a graded capacity, with a degree of understanding depending on observer's familiarity with the observed action. It may be argued that this view is mostly relevant for uncommon actions that require certain training and hence are not in the common human repertoire. This is true but it is consistent with saying that (deep) action understanding refers to either "ordinary" actions common to all people, and to specific actions whose understanding requires certain training in the observer. Hence, we can recognize a wide spectrum of actions, without necessarily understanding them (consistently with our definitions), such as those not in our repertoire. For example, we can recognize that an eagle is circling in the sky, but cannot understand what it means to fly like an eagle in the sky. Since deep understanding and recognition lie on a continuum[17], there exists no dividing line between them.

Our position that action understanding requires (mental) motor simulation is consistent with the view of grounded cognition as introduced by Barsalou (1999). More generally, the phenomenon of understanding, be it an object, an event, or action, is a process that starts in the temporal cortex (because it has first to be perceived) but it soon spreads to frontal areas (and back) resulting in mental simulation. In a wider perspective, understanding something in general draws on frontal areas of the brain. In case of a perceived object located in agent's vicinity, the understanding is enriched by activations of canonical neurons that represent object affordances. In case of a situation, the underlying mental simulation might encode potential reactions of the agent before the agent actually decides for an action.

The idea of mental simulation involving frontal areas is not restricted to understanding behavior but extends itself to language understanding grounded in perception and action. It is known that listening to action verbs (such as *kick* or *pick*) evokes somatotopic activation in the motor cortex (Pulvermüller, 2005). Also on the sentence level, the theory of *language comprehension as motor resonance* is gaining rich empirical support (Zwaan and Taylor, 2006; Glenberg and Kaschak, 2002).

Our view can also be reconciled with rich neuropsychological evidence that points to double dissociations between sensory and motor processes related to action recognition and action execution (Mahon and Caramazza, 2005). The first argument made by Mahon and Caramazza points to the observations that patients impaired in using objects (apraxia) are still able to correctly recognize visually presented actions. And vice versa, patients impaired in recognizing actions associated with the use of objects (pantomime agnosia) are still able to correctly produce object-associated actions. This evidence suggests that perception and production of actions cannot be completely integrated at the level required for correct recognition of an action but some degree of their independence must exist. We agree with this claim which does not conflict with our view, because action recognition takes place in posterior areas (STS) – see section 5.1. As the second argument, testing action (and object) understanding is in Mahon and Caramazza (2005) linked with testing the intactness and usability of conceptual knowledge in patients with various types of sensory or motor impairment. The underlying assumption is that conceptual knowledge cannot be reduced to sensorimotor knowledge. Again, we claim that the sensorimotor simulation (required to allow understanding in the light of embodied theories) is not indispensable in all conceptual tasks, because the correct performance in these experiments can still be achieved by mere recognition of visual features of an action or an object.

7.3 Features of computational models

Although all computational models learn to link perceptual and motor representations, most models are consistent with the "unidirectional" link. The crucial assumption in most models is the reliance on the availability of invariant visual representations generated in STS that are then linked to the motor information. The most simplistic models are based on Hebbian learning between perceptual and motor representations assuming that both are first separately integrated in time and then associated in one-to-one fashion (Keysers and Perrett, 2004). Several models also include the forward model component that provides sensory anticipatory signals to learning associations and add a temporal dimension to both sensory and motor representations (Tani

[17] They can be located at different positions along the depth-of-understanding dimension.

et al., 2004; Oztop et al., 2005; Knott, 2011). Only one model (Tessitore et al., 2010) assigns a direct role of the motor information in interpreting the sensory input, hence putting it in a closer relationship with the direct-matching hypothesis. Only two models, FARS that tackles the canonical part of the MNS and MNS2 extension model, exploit a biologically plausible feedback from the environment in the form of reinforcement learning.

In the context of STS, the visual analysis of actions is a set of complex processes on top of which are STS activations (STS receives inputs both from ventral and dorsal streams). As mentioned above, many computational models assume the existence of viewer-independent object representations, hence taking advantage of the agent-invariant visual patterns used for both executed and observed actions, what greatly simplifies the mirroring task. We argue, however, that the process of acquisition of viewer-independent action representations should be a part of the model. Recently, the machine vision work in action perception has provided candidate approaches for such a task. For instance, Taylor et al. (2007) describe a model, based on a deep belief network (DBN; Hinton et al., 2006) that learns to extract sensible features from moving images in an unsupervised way and another network then associates these preprocessed features with action categories. The functioning of the model seems to resemble well what STS is doing (although the authors do not mention STS at all). Similarly, Chen et al. (2010) present a spatio-temporal extension of the DBN model, tailored to learning invariant perpceptual representations of the motion sequences.

Since the function of the mirror neurons most likely cannot be fully inborn, another important feature of a plausible computational model is to show how the mirroring function develops during experience and how it can be remapped under special circumstances (as shown in the extension of the MNS2 model). We believe that the human ability to understand other minds exploits the mirror neurons and represents a big challenge for computational models, especially when dealing with higher-level phenomena, other than action understanding.

8 Conclusion

This paper offers two contributions. The theoretical one is reconciliationist in spirit. Firstly, we argue that a more satisfactory account of the mirror neuron system must be based on inborn biases as well as include adaptation mechanisms. Secondly, we propose a hypothesis of action understanding as a graded capacity whose manifestation is reflected by the degree of involvement of motor information. The computational part comprises a critical overview of the existing conceptual and computational models of the mirror neuron system. We argue that a plausible computational model of action understanding should account for development and adaptation of the mirroring property, involve graded motor activation that underlies deep action understanding, and include mechanisms for yielding high-level invariant perceptual representations originating in STS, needed for a neural account of action recognition.

Acknowledgement. We thank Alistair Knott, Ľudovít Malinovský, Ľubica Beňušková, and Dana Retová for fruitful discussions. The work was supported by Slovak Grant Agency for Science, #1/0439/11.

References

Alstermark, B., A. Lundberg, U. Norrsell, and E. Sybirska (1981). Integration in descending motor pathways controlling the forelimb in the cat: 9. Differential behavioural defects after spinal cord lesions interrupting defined pathways from higher centres to motoneurones. *Exp. Brain Research 42*, pp. 299–318.

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences 28*(2), pp. 105–24; discussion 125–67.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences 22*(04), pp. 577–660.

Barsalou, L. and K. Wiemer-Hastings (2005). *Situating abstract concepts*, pp. 129–163. Cambridge University Press.

Bekkering, H., A. Wohlschlager, and M. Gattis (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology Section A 53*(1), pp. 153–164.

Benuskova, L. and W. Abraham (2007). STDP rule endowed with the BCM sliding threshold accounts for hippocampal heterosynaptic plasticity. *Journal of Computational Neuroscience 22*(2), pp. 129–133.

Bonaiuto, J. and M. Arbib (2010). Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. *Biological Cybernetics 102*, pp. 341–359.

Bonaiuto, J., E. Rosta, and M. Arbib (2007). Extending the mirror neuron system model, I: Audible actions and invisible grasps. *Biological Cybernetics 96*, pp. 9–38.

Catmur, C., V. Walsh, and C. Heyes (2007). Sensorimotor learning configures the human mirror system. *Current Biology 17*(17), pp. 1527–31.

Chaminade, T., E. Oztop, C. Gordon, and M. Kawato (2008). From self-observation to imitation: visuomotor association on a robotic hand. *Brain Research Bulletin 75*(6), pp. 775–784.

Chen, B., J. Ting, B. Marlin, and N. de Freitas (2010). Deep Learning of Invariant Spatiotemporal Features from Video. *Proceedings of the Workshop on Deep Learning and Unsupervised Feature Learning at the 24th Annual Conference on Neural Information Processing Systems* .

Cohen-Seat, G., H. Gastaut, J. Faure, and G. Heuyer (1954). Etudes expérimentales de lactivité nerveuse pendant la projection cinématographique. *Rev. Int. Filmologie 5*, pp. 7–64.

Craighero, L., A. Bello, L. Fadiga, and G. Rizzolatti (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia 40*(5), pp. 492–502.

Cuijpers, R., F. Stuijt, and I. Sprinkhuizen-Kuyper (2009). Generalisation of action sequences in RNNPB networks with mirror properties. In *Proceedings of the 17th European Symposium on Artificial Neural Networks (ESANN 2009)*, Bruges, Belgium, pp. 251–256.

Elman, J., E. Bates, A. Johnson, A. Karmiloff-Smith, D. Parisi, and D. Plunkett (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Fadiga, L., L. Fogassi, G. Pavesi, and G. Rizzolatti (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology 73*(6), pp. 2608.

Fagg, A. (1996). *A Computational Model of The Cortical Mechanisms Involved in Primate Grasping.* Ph. D. thesis, University of Southern California, Computer Science Department.

Fagg, A. and M. Arbib (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks 11*, pp. 1277–1303.

Farivar, R. (2009). Dorsal-ventral integration in object recognition. *Brain Research Reviews 61*(2), pp. 144–153.

Frischen, A., D. Loach, and S. Tipper (2009). Seeing the world through another persons eyes: Simulating selective attention via action observation. *Cognition 111*, pp. 212–218.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks 16*, pp. 1325–1352.

Froemke, R. and Y. Dan (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature 416*(6879), pp. 433–438.

Gallese, V., L. Fadiga, L. Fogassi, and G. Rizzolatti (1996). Action recognition in the premotor cortex. *Brain: A Journal of Neurology 119*, pp. 593–609.

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

Gastaut, H. and J. Bert (1954). EEG changes during cinematographic presentation; moving picture activation of the EEG. *Electroencephalography and Clinical Neurophysiology 6*(3), pp. 433.

Gazzola, V., H. van der Worp, T. Mulder, B. Wicker, G. Rizzolatti, and C. Keysers (2007). Aplasics born without hands mirror the goal of hand actions with their feet. *Current biology 17*(14), pp. 1235–1240.

Gibson, J. (1977). The theory of affordances. *Perceiving, acting, and knowing: Toward an ecological psychology* , pp. 67–82.

Glenberg, A. and M. Kaschak (2002). Grounding language in action. *Psychonomic Bulletin & Review 9*(3), pp. 558–565.

Goldman, A. (2008). Does one size fit all? Hurley on shard circuits. *Behavioral and Brain Sciences 31*, pp. 27–28.

Grezes, J., J. Armony, J. Rowe, and R. Passingham (2003). Activations related to mirror and canonical neurones in the human brain: an fMRI study. *Neuroimage 18*(4), pp. 928–937.

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

Hasselmo, M., E. Rolls, G. Baylis, and V. Nalwa (1989). Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research 75*, pp. 417–429.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences 5*(6), pp. 253–261.

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews 34*(4), pp. 575–83.

Hickok, G. (2008). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience 21*(7), pp. 1229–43.

Hickok, G. and M. Hauser (2010). (Mis)understanding mirror neurons. *Current Biology 20*(14), pp. R593–4.

Hinton, G., S. Osindero, and Y. Teh (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation 18*(7), pp. 1527–1554.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences 79*, pp. 2554–2558.

Hurley, S. (2008). The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behavioral and Brain Sciences 31*(01), pp. 1–22.

Iacoboni, M. (2009). Imitation, empathy and mirror neurons. *Annual Review of Psychology 60*, pp. 653–670.

Iacoboni, M., L. Koski, M. Brass, H. Bekkering, R. Woods, M. Dubeau, J. Mazziotta, and G. Rizzolatti (2001). Re-afferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences 98*, pp. 13995–9.

Ishida, H., K. Nakajima, M. Inase, and A. Murata (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *Journal of Cognitive Neuroscience 22*(1), pp. 83–96.

Izhikevich, E. and N. Desai (2003). Relating STDP to BCM. *Neural Computation 15*(7), pp. 1511–1523.

James, W. (1890). *The Principles of Psychology (Vols. 1 & 2).* New York: Henry Holt and Company.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology 9*(6), pp. 718–727.

Keysers, C. and D. Perrett (2004). Demystifying social cognition: a Hebbian perspective. *Trends in Cognitive Sciences 8*(11), pp. 501–507.

Kilner, J., K. Friston, and C. Frith (2007). The mirror-neuron system: a Bayesian perspective. *NeuroReport 18*(6), pp. 619–623.

Knoblich, G. and R. Flach (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science 12*(6), pp. 467.

Knott, A. (2011). *Sensorimotor Cognition and Natural Language Syntax.* The MIT Press.

Kraskov, A., N. Dancause, M. Quallo, S. Shepherd, and R. Lemon (2009). Corticospinal neurons in macaque ventral premotor cortex with mirror properties: a potential mechanism for action suppression? *Neuron 64*(6), pp. 922–930.

Mahon, B. and A. Caramazza (2005). The orchestration of the sensory-motor systems: Clues from neuropsychology. *Cognitive Neuropsychology 22*(3/4), pp. 480–494.

Mason, C., J. Gomez, and T. Ebner (2001). Hand synergies during reach-to-grasp. *Journal of Neurophysiology 86*(6), pp. 2896–2910.

Matelli, M. and G. Lupino (2001). Parietofrontal circuits for action and space perception in the macaque monkey. *NeuroImage 14*, pp. 27–32.

Mel, B. and C. Koch (1990). Sigma-pi learning: on radial basis functions and cortical associative learning. In D. Touretzky (Ed.), *Advances in NIPS-2*, pp. 474–481. The MIT Press.

Miall, M. (2003). Connecting mirror neurons and forward models. *NeuroReport 14*(17), pp. 2135–2137.

Miikkulainen, R. and M. Dyer (1988). Forming global representations with extended backpropagation. Technical Report CSD-880039, Computer Science Department, University of California at Los Angeles.

Milner, A. and M. Goodale (1995). *The Visual Brain in Action.* Oxford: Oxford University Press.

Mukamel, R., A. Ekstrom, J. Kaplan, M. Iacoboni, and I. Fried (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology 20*(8), pp. R353–R354.

Oberman, L. and V. Ramachandran (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin 133*(2), pp. 310.

Oberman, L. and V. Ramachandran (2008). How do shared circuits develop? *Behavioral and Brain Sciences 31*, pp. 34–35.

Olson, C. (2003). Brain representation of object-centred space in monkeys and humans. *Annual Review of Neuroscience 26*, pp. 331–354.

O'Reilly, R. and Y. Munakata (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain.* The MIT Press.

Oztop, E. and M. Arbib (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics 87*, pp. 116140.

Oztop, E., M. Kawato, and M. Arbib (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks 19*(3), pp. 254–271.

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

Peeters, R., L. Simone, K. Nelissen, M. Fabbri-Destro, W. Vanduffel, G. Rizzolatti, and G. Orban (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *Journal of Neuroscience 29*(37), pp. 11523.

Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research 91*(1), pp. 176–180.

Pelphrey, K., J. Morris, and G. Mccarthy (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience 16*(10), pp. 1706–1716.

Perrett, D., M. Oram, M. Harries, R. Bevan, J. Hietanen, P. Benson, and S. Thomas (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research 86*(1), pp. 159–73.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience 6*(7), pp. 576–582.

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

Rizzolatti, G., R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research 71*(3), pp. 491–507.

Rizzolatti, G. and L. Craighero (2004). The mirror-neuron system. *Annual Review of Neuroscience 27*, pp. 169–92.

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

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

Rizzolatti, G. and C. Sinigaglia (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience 11*(4), pp. 264–74.

Rumelhart, D., J. McClelland, and the PDP research group (1986). *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Volume 1. Cambridge, MA: MIT Press.

Shepherd, S., J. Klein, R. Deaner, and M. Platt (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences 106*(23), pp. 9489.

Spivey, M. (2007). *The Continuity of Mind.* Oxford: Oxford University Press.

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

Taylor, G., G. Hinton, and S. Roweis (2007). Modeling human motion using binary latent variables. *Advances in Neural Information Processing Systems 19*, pp. 1345.

Tessitore, G., R. Prevete, E. Catanzariti, and G. Tamburrini (2010). From motor to sensory processing in mirror neuron computational modelling. *Biological Cybernetics 103*, pp. 471–485.

Tkach, D., J. Reimer, and N. G. Hatsopoulos (2007). Congruent activity during action and action observation in motor cortex. *The Journal of Neuroscience 27*(48), pp. 13241–50.

Umiltà, M. A., L. Escola, I. Intskirveli, F. Grammont, M. Rochat, F. Caruana, a Jezzini, V. Gallese, and G. Rizzolatti (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences 105*(6), pp. 2209–13.

Ungerleider, L., M. Mishkin, et al. (1982). Two cortical visual systems. *Analysis of visual behavior 549*, pp. 586.

van Rooij, I., W. Haselager, and H. Bekkering (2008). Goals are not implied by actions, but inferred from actions and contexts. *Behavioral and Brain Sciences 31*, pp. 38–39.

Wiedermann, J. (2003). Mirror neurons, embodied cognitive agents and imitation learning. *Computing and Informatics 22*(6), pp. 545–559.

Wolpert, D., K. Doya, and M. Kawato (2003). A unifying computational framework for motor control and social interaction. *Philos. Trans. Royal Soc. London B 358*, pp. 593–602.

Zwaan, R. and L. Taylor (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Exp. Psychology: General 135*, pp. 1–11.